

# ESSAYS IN THE ECONOMICS OF EDUCATION

by

**Yang Song**

B.A. International Economics and Trade,  
Renmin University of China, 2010

M.A. Economics, University of Pittsburgh, 2011

Submitted to the Graduate Faculty of  
the Kenneth P. Dietrich School of Arts and Sciences in partial  
fulfillment

of the requirements for the degree of

**Doctor of Philosophy**

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH  
KENNETH P. DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Yang Song

It was defended on

April 13th, 2015

and approved by

Dr. Daniel Berkowitz, University of Pittsburgh

Dr. George Loewenstein, Carnegie Mellon University

Dr. Thomas Rawski, University of Pittsburgh

Dr. Werner Troesken, University of Pittsburgh

Dissertation Director: Dr. Daniel Berkowitz, University of Pittsburgh

Copyright © by Yang Song  
2015

## ESSAYS IN THE ECONOMICS OF EDUCATION

Yang Song, PhD

University of Pittsburgh, 2015

This dissertation comprises three essays in the economics of education. I use natural and field experiments to evaluate education policies and programs. I also bring insights from psychology to understand how to improve learning and work efficiency.

The first chapter shows how a Chinese city was successful in helping its low-performing schools catch up. The city's education bureau identified several low-performing middle schools and guaranteed elite high school admission to their top ten-percent graduates. I document that schools affected by this top-ten percent policy improved their performance by 0.3 standard deviation. To understand the underlying mechanisms, the city's lottery system for middle school assignment is used to test for changes in composition and value-added. The study suggests that incentives for better students to attend lower-performing schools help narrow not only the school performance gap but also the school quality gap.

The second chapter evaluates a peer mentoring program that matches high-performing students as mentors to their low-performing classmates and provides non-monetary incentives for them to study together and stay in school. We implemented the program in two rural Chinese middle schools. The program did not improve the mentees' math scores, but instead increased their learning stress. However, the program did significantly improve the mentors' math scores by 0.57 standard deviations and lowered their dropout rate by 3%, with no impact on their mental health scores. We discuss possible reasons for these

surprising results and propose changes in program design that may help mentees benefit as well.

The third chapter studies the effect of time abundance on work efficiency. I propose a strategic framework of efficient completion of time-constrained tasks. Facing a task with a deadline, an agent is under-motivated when there is abundant time and over-motivated when the deadline is too close. This generates a hump-shaped relationship between efficiency and time available for the task. I use online homework tracking data for a large introductory microeconomics class to test this theory. Within-subject analysis provides evidence supporting the predictions: when a student starts work neither too early nor too late, he/she has a higher class ranking and a lower time cost.

## TABLE OF CONTENTS

<b>1.0 SORTING, SCHOOL PERFORMANCE AND SCHOOL QUALITY: EVIDENCE FROM CHINA</b>	<b>1</b>
1.1 Introduction	1
1.2 Policy Backgroundd	6
1.2.1 Top Ten-Percent Quota Policy	6
1.2.2 Preference-based Lottery Middle School Assignment	8
1.3 Data	9
1.4 Does the Top Ten-Percent Quota Policy work?	15
1.4.1 Comparisons of Trends in Performance	15
1.4.2 Test for Selection in Treatment Status	18
1.4.3 Difference-in-Differences Estimation	20
1.4.4 Placebo Test	24
1.4.5 Change in Distributions of Academic Performances	24
1.5 Underlying Mechanisms	26
1.5.1 A Conceptual Framework	26
1.5.2 Using Lottery Records to Tease Out Mechanisms	28
1.5.3 Change in Composition	29
1.5.4 Change in Value-added	34
1.6 Conclusion	45

<b>2.0 PEER MENTORING AND GROUP INCENTIVES: EVIDENCE FROM CHINESE RURAL MIDDLE SCHOOLS</b>	47
2.1 Introduction	47
2.2 Program Design	50
2.2.1 Pairing method	50
2.2.2 Group Incentive Design	51
2.2.3 Implementation	52
2.2.4 Balance Check: Pre-Treatment Characteristics	53
2.3 Evaluating the Peer Mentoring Program	54
2.3.1 Main Results	56
2.3.2 Treatment Effects on Subcategories of Mental Health Survey, Study Attitude, Effort and Behavior	57
2.3.3 Heterogeneous Treatment Effects	64
2.3.4 Subgroup Treatment Effects	69
2.3.5 Qualitative Measures	70
2.4 Conclusions	71
<b>3.0 EFFICIENT PROCRASTINATION</b>	74
3.1 Introduction	74
3.2 Conceptual Framework	75
3.3 Data	77
3.3.1 CourseWeb Tracking Data	77
3.3.2 Measurements of Procrastination, Time Cost and Performance	78
3.4 Empirical Methodology and Results	83
3.4.1 Between-subject Analysis	83
3.4.2 Within-subject Analysis	88
3.5 Conclusion and Future Work	95
<b>APPENDIX.</b>	97

A.1	Appendix for Chapter 1 . . . . .	97
A.2	Appendix for Chapter 2 . . . . .	103
<b>BIBLIOGRAPHY</b>	. . . . .	108



## LIST OF TABLES

1.1	Administrative Panel Data Description . . . . .	13
1.2	Individual Level Data: Summary Statistics 2007 . . . . .	14
1.3	Test for Treatment Status Selection . . . . .	19
1.4	Difference-in-differences: Treatment Effect on School Characteristics . . . . .	21
1.5	Incremental Treatment Effect on School Characteristics . . . . .	23
1.6	Composition effect Change in Percentages of Sixth Graders Choosing Policy Schools . . . . .	31
1.7	Are elementary students with high ability and high SES more likely to choose a policy school after the policy? . . . . .	32
1.8	Conditional logit estimates of choosing a policy school: Who Are Switching? . . . . .	33
1.9	Lottery Randomness Verification . . . . .	38
1.10	First Stage: Use losing a lottery to instrument for policy school attendance . . . . .	40
1.11	Average Change in School Quality Gap: 2SLS Results . . . . .	41
1.12	Change in Distributions of School Quality Gap: IV Quantile Treatment Ef- fect . . . . .	43
2.1	Baseline Student Characteristics and Balance Check . . . . .	55
2.2	Treatment Effect on Math, Overall Mental Health and Dropout Probability . . . . .	58
2.3	Treatment Effect on other variables of possible interests Mentee and Placebo Mentee Subsample . . . . .	60

2.4	Treatment Effect on other variables of possible interests Mentor and Placebo	
	Mentor Subsample . . . . .	61
2.5	Heterogeneous Effect on Math Score: Quantile Regressions . . . . .	68
2.6	Heterogeneous Effect on Mental Health Score: Quantile Regressions . . . . .	69
2.7	Student Subjective Survey Evaluation . . . . .	72
3.1	Summary Statistics . . . . .	79
3.2	Between-subject analysis: Performance and Timing . . . . .	87
3.3	Between-subject analysis: Time cost (total entries) and timing . . . . .	89
3.4	Within-Subject Analysis: Normalized Timing and Homework Performance . . . . .	91
3.5	Within-Subject Analysis: Timing and Time Cost (total entries) . . . . .	92
A1	Description of Data Sets . . . . .	99
A2	Individual Level Data: Summary Statistics 2005 . . . . .	100
A3	Individual Level Data: Summary Statistics 2006 . . . . .	101
A4	Falsification Test for Treatment Effects . . . . .	102
A5	Balance Check for the Whole Sample . . . . .	103
A6	Balance Check for Mentor/Mentee Subsamples . . . . .	104
A7	Main Results using the Pooled Sample . . . . .	105
A8	Heterogeneous Treatment Effect for Different Grade and Gender . . . . .	106
A9	Student Suggestions . . . . .	107

## LIST OF FIGURES

1.1	Timeline of Policy Announcement and Data Availability for Different Cohorts	10
1.2	Normalized 9th Grade Scores: Policy Schools and Other Schools . . . . .	16
1.3	General Pattern of Changes in Score Distributions . . . . .	25
1.4	Quantile Regression Estimates . . . . .	26
1.5	Changes in Value-added Gaps . . . . .	44
2.1	Kernel Density Plots of Mentees' Standardized Math Scores . . . . .	62
2.2	Kernel Density Plots of Mentors' Standardized Math Scores . . . . .	63
2.3	Kernel Density Plots of Mental Health Scores . . . . .	64
2.4	Quantile Treatment Effects on Standardized Math Scores . . . . .	66
2.5	Quantile Treatment Effects on Mental Health Scores . . . . .	67
3.1	Homework Start/Submit Timing Patterns . . . . .	81
3.2	Homework Start/Submit Timing Patterns . . . . .	82
3.3	Between-Subject Analysis: Relationship between Start/Do/Submit Early and Homework Performance . . . . .	85
3.4	Between-Subject Analysis: Relationship between Start/Do/Submit Early and Time Cost . . . . .	86
3.5	Within-Subject Analysis: Relationship between Start/Do/Submit Early and Homework Performance . . . . .	93

3.6 Within-Subject Analysis: Relationship between Start/Do/Submit Early and Time Cost . . . . .	94
--	----

## Preface

This work would not have been possible without the advice and support by Daniel Berkowitz, George Loewenstein, Thomas Rawski, Werner Troesken and Randall Walsh. I also wish to thank Saurabh Bhargava, Eric Chyn, Susan Dynarski, Sebastian Galiani, Marc Law, Ian Morrall, M. Najeeb Shafiq, Tate Twinam, Lixin Colin Xu, as well as participants of CIERS at the University of Michigan, 2014 Xiamen Ronald Coase Institute Workshop, 2nd Summer School of Socioeconomic Inequality at the University of Chicago, Applied Micro Seminar and Development Seminar at the University of Pittsburgh, and CBDR Brownbag at Carnegie Mellon University for their helpful comments.

I also wish to thank the Changsha education bureau for their cooperation and help with data collection. For the second chapter, my coauthors and I would like to thank Chu Yang for invaluable input at an early stage of the project. We also thank Fei He and research assistants at CEEE, Shaanxi Normal University for their help with implementation. We acknowledge the financial support from of the Chinese National Natural Science Foundation of China (grant numbers 71033003 and 71333012). For the third chapter, I thank Shirley Cassing, Brian Beach, Lise Vesterlund, Stephanie Wang, and Nicholas Landolina for their help and suggestions with data collection and IRB approval. I am grateful for the financial support from the Chou Fellowship, Andrew Mellon Predoctoral Fellowship, and travel grants from the Economics Department, the Dietrich School of Arts and Sciences and Asian Studies Center at the University of Pittsburgh.

For their emotional and material support, my final thanks are to my parents.

## 1.0 SORTING, SCHOOL PERFORMANCE AND SCHOOL QUALITY: EVIDENCE FROM CHINA

### 1.1 INTRODUCTION

School choice policies including vouchers and lotteries have been widely adopted in many countries. The underlying idea is that these policies give students and parents more freedom in choosing schools, and schools are under more pressure to improve quality to attract students. However, these school-choice policies have had mixed results and, somewhat problematically, have in some cases increased sorting.<sup>1</sup> When given choices, students with low socioeconomic status (SES) are less likely to choose a good school ([Ajayi, 2011](#); [Walters, 2013](#)).<sup>2</sup> Theoretical and empirical evidence shows that students with high SES and high ability sort out of low-performing schools, leaving disadvantaged students behind ([Epple and Romano, 1998](#); [Levin, 1998](#); [Hsieh and Urquiola, 2006](#); [Galiani et al., 2008](#); [Chakrabarti, 2009](#)).<sup>3</sup> Given the importance of peer composition in the education production function

---

<sup>1</sup>See [Hoxby \(2000\)](#); [Bettinger \(2005\)](#); [Rothstein \(2007\)](#); [Chakrabarti \(2008\)](#) for mixed results on effects of introducing school choice and increased school competition; see [Rouse and Barrow \(2009\)](#); [MacLeod and Urquiola \(n.d.\)](#) for reviews. A large literature on the impacts of winning a school lottery or school voucher have also found mixed results in various locations, like Milwaukee, Columbia, New York, Chicago, Charlotte-Mecklenburg and China. See [Abdulkadiroğlu et al. \(2011\)](#); [Angrist et al. \(2002\)](#); [Dobbie and Fryer \(2011\)](#); [Angrist et al. \(2013b\)](#); [Cullen et al. \(2006\)](#); [Deming \(2011\)](#); [Deming et al. \(2014\)](#); [Krueger and Zhu \(2004\)](#); [Peterson et al. \(1998\)](#); [Rouse \(1998\)](#); [Rouse and Barrow \(2009\)](#); [Witte \(1997\)](#); [Zhang \(2012\)](#).

<sup>2</sup>See [Hoxby and Avery \(2013\)](#) for similar patterns on selective college applications.

<sup>3</sup>[Levin \(1998\)](#) reviews empirical evidence on voucher programs and find consistent results that school choice leads to greater SES and racial segregation. See [Hoxby and Avery \(2013\)](#) for similar patterns on selective college applications. [Muralidharan and Sundararaman \(2013\)](#) found no negative spillover effect of private school vouchers on students staying in public schools.

(Epple and Romano, 2011), this sorting pattern may lead to more inequality (Epple and Romano, 1998; Calabrese et al., 2012). How to change sorting given choices so as to close the performance gaps is still an open question.

This study evaluates a policy that provides incentive for high ability students to voluntarily enroll in low-performing schools under a choice-based lottery school assignment system. Changsha, a Chinese provincial capital city with a population of seven million, introduced the top ten-percent quota policy in 2007. The education bureau chose one or two low-performing public middle schools in each district and guaranteed admission to an elite high school for the top 10% of 9th grade graduates from each of these schools. This paper answers two questions: Did the Top 10% Quota Policy narrow the school performance gap? And if so, what are the underlying mechanisms?

To estimate the policy impact, I employ a difference-in-differences identification strategy with a panel data set of middle school graduation exam performance from 2004 to 2011. I show that the policy schools improved their average performances by around 0.3 standard deviations in the middle school graduation exam and increased their elite high school attendance rates by around six percent.

This impact could be working through various possible mechanisms, including a composition effect, a tournament effect, and a peer effect. First, a conceptual framework in section 5 illustrates the trade-offs on whether to change enrollment choice from an over-subscribed school to a low-performing policy school. It predicts that above average students who are not at the top of the talent distribution are most likely to switch to low-performing policy schools.<sup>4</sup> Second, competition to place at the top ten-percent may stimulate a higher effort level exerted by students, especially the top-performing ones. Third, with better peer groups and a more active learning environment, it may bring positive spill-over effects on non-switchers: students who chose an over-subscribed school but lost the lottery and

---

<sup>4</sup>Benefit from switching occurs when a student has a higher probability of making it into the top ten-percent of a low-performing school than that of making it into the top thirty-percent among all students in the city; cost is having lower quality peers.

students who would have chosen these low-performing schools anyway. The first channel redistribute students across schools; the latter two channels increase the value-added at the policy schools.

To tease out the mechanisms, I exploit Changsha’s preference-based lottery middle school assignment. Since 1996, the Changsha education bureau has assigned a fixed number of seats in several neighborhood public middle schools to each elementary school every year. A sixth grade student chooses one from the short list of middle schools assigned to his/her elementary school.<sup>5</sup> In cases of over-enrollment, a lottery takes place and randomly assigns winners to their chosen school and losers to the under-subscribed low-performing school, some of which were assigned the top ten-percent quotas. This allows me to analyze changes in students’ school choices and compare the outcomes of lottery winners with those of lottery losers to obtain unambiguous results on the value-added gap between policy schools and over-subscribed schools.

Observing the school choices by sixth graders, I compare the baseline performance of students who voluntarily chose the policy schools before and after the policy. I found that sixth graders with high math scores and high SES were more likely to choose a policy school after the policy. In particular, students with high, but not the highest, sixth grade math scores changed their school choice to policy schools after the policy introduction. This result is consistent with predictions from the conceptual framework.

Using lottery assignment as the instrument variable, I estimate the local average treatment effect (LATE) of attending a policy school before and after the introduction of the top ten-percent quota policy. Estimates show that policy school attendance caused a 0.3 standard deviation decrease in academic performance of lottery losers before the policy; this value-added gap was closed after the policy. With better peer quality, policy schools may improve their value-added for all students; extra effort to place at the top ten-percent brings a tournament effect only for the high-performing students. To estimate heteroge-

---

<sup>5</sup>In China, elementary school goes from first to sixth grade and middle school goes from seventh to ninth grade.



neous effects, I conduct instrumental variable quantile treatment effect (QTE) analysis (Abadie et al., 2002) to test how the policy changed the distribution of value-added gap. Results shows that before the policy, the value-added gap was negative for most deciles across the distribution, and more so for high-performing students. After the policy, the value-added gaps were closed for most of the deciles, except for the sixtieth and ninetieth percentile. Since students at low quantiles would only be subject to changes in peer effects but not tournament effects, improvements on value-added at low deciles suggest that peer effects are at work. The policy closed not only the performance gap but also the value-added gap between the low-performing policy schools and the over-subscribed schools.

This study has some implications on recent school choice reform. Attending private or charter schools sometimes brings academic and/or nonacademic benefits to the lottery winners.<sup>6</sup> However, school choice reforms may increase sorting and may widen the performance gap (Epple and Romano, 1998; Levin, 1998; Hsieh and Urquiola, 2006; Calabrese et al., 2012; Galiani et al., 2008; Chakrabarti, 2009). Similar with previous studies (Ajayi, 2011; Butler et al., 2013; Hastings et al., 2008; Hoxby and Avery, 2013; Walters, 2013), I find that students with lower SES are less likely to choose an over-subscribed school. Previous research attempted to improve school choice by providing information for parents and students and found that it helps in some cases, but not in others.<sup>7</sup>

Aside from efforts to help disadvantaged students choose and attend better schools, many studies have also looked more directly on how to improve the quality of low-performing schools. Angrist et al. (2013a) and Dobbie and Fryer (2013) found that the “No Excuses” model of urban education is the key to charter school effectiveness. In the case of the top ten-percent quota policy, combining school choice with incentives for good students to

---

<sup>6</sup>A large literature has looked at the effect of attending a chosen school in a lottery setting and has found mixed evidence (Abdulkadiroğlu et al., 2011; Angrist et al., 2002; Dobbie and Fryer, 2011; Angrist et al., 2013b; Cullen et al., 2006; Deming, 2011; Deming et al., 2014; Krueger and Zhu, 2004; Peterson et al., 1998; Rouse, 1998; Rouse and Barrow, 2009; Witte, 1997).

<sup>7</sup>Positive effects of providing information were found in Chicago and Pakistan (Hastings and Weinstein, 2008; Andrabi et al., 2009), but not in India or Chile (Banerjee et al., 2010; Mizala and Urquiola, 2013).

attend lower-performing schools helps change the sorting patterns and narrow the performance gap. More importantly, instrumental QTE results show that the value-added gap is also closed almost everywhere across the distribution, which suggests that positive peer effects brought on by the composition changes are at work.

This study also relates to how relative evaluation changes sorting. A similar policy in the U.S. is the top x-percent rule in Texas, California and Florida, which guarantees flagship state university admission for top x-percent of seniors in all high schools. These policies and Changsha’s policy differ in the school level (middle school v.s. high school), in the affected schools (some low-performing schools v.s. all schools) and most importantly, in their purposes. Changsha introduced the top ten-percent quota policy to improve low-performing schools by changing composition and improving school quality collectively, while top x% in the U.S. mainly aims at improving the minority students representation in selective colleges after the affirmative action ban (Long, 2004; Long et al., 2010). Therefore, while sorting emerges in both cases, change in sorting was unintended in Texas (Cullen et al., 2013), but it was expected and beneficial in the case of Changsha.

Results here on the top ten-percent quota policy complement previous findings on the Texas Ten-Percent Law. Cullen et al. (2013) and this study provide converging evidence that relative evaluation brings different sorting behaviors and improves student composition in low-performing schools. Although the top x-percent rule fails to promote the opportunity of minority groups as well as the affirmative action (Long, 2004; Long et al., 2010), it helps low-performing high schools to improve their performance faster than other schools (Cortes and Zhang, 2011). I also find that in the case of Changsha, policy schools caught up in their performance. Further, this study advances previous studies by estimating the value-added gap before and after the policy, exploiting Changsha’s unique lottery school assignment. Results show that the school quality gap was closed as well. To what extent these results would apply to environments with large variations in racial composition and instructions is unknown.

The rest of the paper is organized as follows. Section 2 provides the background on the top ten-percent quota policy and choice-based lottery middle school assignment. Section 3 describes data construction, and thus why and how I use the data to conduct the analysis. Section 4 evaluates the policy impact on school performances using a difference-in-differences framework. Section 5 provides a conceptual framework on school choice and exploits the preference-based lottery middle school assignment to tease out changes in composition and in value-added. Section 6 concludes.

## **1.2 POLICY BACKGROUND**

Although China has experienced rapid economic growth in the past few decades, enlarging inequality has brought pressing social problems. School choice is an especially controversial topic. Differences in education quality and academic performance across schools have been widened, both within and across cities. Students bear heavy pressure to compete for access to good schools and avoid bad schools, starting from a very early age. Lump-sum fees for high quality schools and an increasingly large and expensive tutoring industry put students with low SES in worse situations. To alleviate these problems and improve equal education opportunities, governments from the central to local level have been implementing various policies to equalize school quality and education opportunities for students. One of such policies is Changsha’s top ten-percent quota policy.

### **1.2.1 Top Ten-Percent Quota Policy**

Changsha is a provincial capital city in South-central China, with a population of about seven million. While there are several rural districts/counties in the city, the top ten-

percent quota policy only relates to the five urban districts.<sup>8</sup> At the elementary school level in these urban districts, there are around 18,000 students in each cohort and about 240 schools; at the middle school level, there are around 20,000 students in each cohort and about 75 schools. Elementary schools run from first until sixth grade, middle schools run from seventh until ninth grade, and high schools run from tenth until twelfth grade. At the end of their ninth grade, students take the Middle School Graduation Exam (MSGEx), which determines high school admissions.

Middle schools with better past MSGEx performances carry better reputations of school quality and attract students with better academic performance and higher socioeconomic status. Large performance gaps intensify sorting by ability across schools. In 2007, Changsha’s education bureau initiated and announced the top ten-percent quota policy.<sup>9</sup> One or two low-performing middle schools in each district were chosen to pair with an elite high school. Six middle schools were originally assigned the quota since 2007 and five more were added in 2008. Altogether, these 11 policy schools have around 3,000 students per cohort, about 14% of the total middle school student population in the city. As of the writing of this paper, the policy is still operating.

For each paired group of a low-performing middle school and an elite high school, the top Ten-Percent 9th grade graduates from the middle school every year are guaranteed to be admitted into the elite high school, without taking the MSGEx at the end of 9th grade and competing with all other graduates. Although the top ten-percent ranking method is decided by individual middle schools and varies slightly, they all basically use accumulated performance across subjects throughout the three years in middle school. To be eligible to compete for the top ten-percent, students are required to be admitted through the preference-based lottery and attend the school from 7th grade onward. These requirements rule out possibility of late-term transfers.

---

<sup>8</sup>The district description of urban and rural here is from a Chinese perspective. Urban areas are more developed and populated and typically have better schools.

<sup>9</sup>The project is called “dui kou zhi sheng” in Chinese pinyin, which literally translates to “pair-wise direct admission”.

The main goals of the Top Ten-Percent Policy are to change sorting, narrow performance gap between middle schools and provide better education for students attending the lower-performing middle schools. More equalized performance across middle schools eases the concern of parents and lower the incentive of sorting. Parents would not worry as much if they send their children to a slightly lower-performing school since peer quality and chances to attend a good high school would now be higher. Students from low SES families who attend a lower-performing school would still get comparable value-added during middle school.

### **1.2.2 Preference-based Lottery Middle School Assignment**

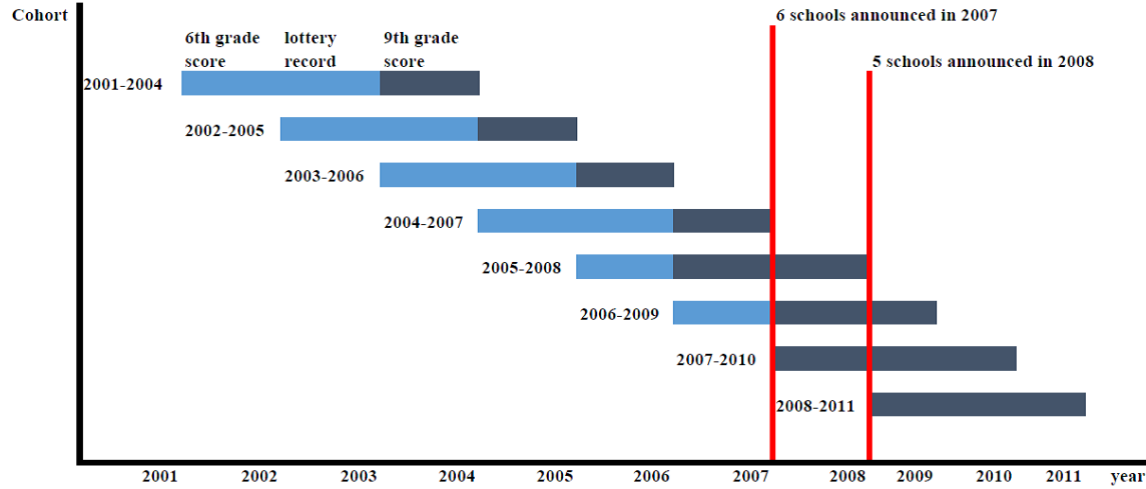
The unique preference-based lottery middle school assignment in Changsha allows me to tease out the mechanisms. Since 1996, Changsha introduced the preference-based lottery middle school assignment. Each elementary school is assigned a fixed number of seats in two or three neighboring middle schools for its graduates (i.e. sixth graders). Each sixth grader can only choose one middle school. If a middle school is over-enrolled from a particular elementary school, a lottery takes place and assigns winners to the chosen middle school and losers to a school that has unfilled seats for that elementary school. Lottery losers will be assigned randomly to one of the under-subscribed middle schools if there are more than one of them.

Before the lottery, part of students get pre-admitted to several designated schools with specialized training in art, music, dance, athletics, or foreign languages. Official rules forbid other middle schools to pre-admit students by organizing selection exams. They also require all students to obey the preference-based lottery middle school assignment. I find that students who were pre-admitted or participated in the lottery and chose a high-reputation school have better family background and better academic performance than those who chose a low-performing school.

### 1.3 DATA

Through the generous cooperation of local officials, I was allowed access to restrictive administrative data sets. [Figure 1.1](#) illustrates the time line of policy introduction and data availability for different years. Individual level 9th grade Middle School Graduation Exam (MSGGE) scores were available for eight cohorts, who entered into middle school through 2001 to 2008 and graduated three years later through 2004 to 2011. For the latter four cohorts, I have the lottery middle school assignment records, which happened in 2005 through 2008. In addition, I collected elementary school graduation exam scores from one school district for the last two cohorts. I also observe students' gender, ethnicity, city residency (hukou) and parental political affiliation for the last two cohorts.

Figure 1.1: Timeline of Policy Announcement and Data Availability for Different Cohorts



Notes: Data is available on the eight cohorts (2001-2004 cohort means they graduated from elementary schools in 2001 as a sixth grader and finished middle school in 2004 three years later as a ninth grader). For each cohort, the colors of each bar indicate data availability of three records: light blue for “not available” and dark blue for “available”; from left to right these three bars refer to 6th grade score, lottery record and 9th grade score. The red vertical line indicates that six and five policy schools were announced in 2007 and 2008 respectively.

Only observing sixth grade baseline scores for 2007 and 2008 in Yuhua district is the reason why I conduct analysis on quality change in composition for the 2008 policy school in this district. To verify the representativeness of the results, I ran regressions of average performance as the dependent variable on interaction between dummies for policy and for Yuhua district. The coefficient for interaction term is insignificant, which suggests that comparison between policy schools and other schools in Yuhua district is similar with that in other districts. In a separate regression, coefficient for the triple interaction of three dummies (policy, Yuhua district and post-policy) is also insignificant, which indicates that

the policy treatment effect is no different in Yuhua district than the other districts.<sup>10</sup>

Looking at students' school choice reveals sorting patterns. Table 1.2 shows the summary statistics for 2007-2010 cohort. Pre-admitted students have significantly higher baseline scores and better socioeconomic status than those going through preference-based lottery school assignment. Among those students going through the normal procedure of the preference-based lottery school assignment, students choosing over-subscribed schools have higher baseline scores and better socioeconomic status. This echoes with recent literature showing that students from low socioeconomic background are unlikely to choose high-quality schools across many settings (Ajayi (2011); Butler et al. (2013); Hastings et al. (2008), Hoxby and Avery (2013)).

The main outcome variables in this paper are the 9th grade MSGE scores and elite high school attendance. The exam is high-stake since its score is the only criteria for high schools to select students<sup>11</sup>. Notice that after the policy, the top ten-percent of students in the policy impacted schools get direct admission to the elite high school without taking the MSGE.

The MSGE final grade is consist of 6 parts, including Chinese, math, English, social science (history and politics), science (physics and chemistry), integrated subjects (biology, geography, physical education). In 2004 and 2005, final grade were in scores and high schools admissions followed a clear score cutoff. Since 2006, the education bureau changed from actual scores to letter grades A-E, with A being the highest grade and E the lowest.<sup>12</sup> The letter grades are determined by the percentile of student performance in each subject: top 25% gets an A, the next 35%, 20% and 10% gets a B,C and D respectively, and the

---

<sup>10</sup>Results are available upon request.

<sup>11</sup>Few exceptional students get directly admitted because of the quota policy, athletic or music specialties, or exceptional academic excellence. Proportion of students who get directly admitted through channels other than quota policy did not change.

<sup>12</sup>The transition in grading scale might be the reason why 9th grade scores in 2006 are significantly lower than other years. Since the grading scale change equally influences students in policy schools and other schools, it does not affect the analysis of this study. I tested the effect of grading scale change on students' relative ranking in the city using 2004 and 2005 data and found that students in policy schools do not experience different change in ranking than other schools.



bottom 10% gets a failing grade E. High schools admit students based on letter grades and prefer higher and more balanced grades. For example, four A's and two B's is preferred to five A's and one C. In the analysis, letter grades A to E are treated as the average percentile of that grade, i.e. 0.875, 0.575, 0.275, 0.1, 0.05. The highest score is  $0.875 \times 6 = 5.25$  (all 6 subjects are As) and lowest is 0.3. To make the grades comparable before and after the grading scale change, I assign letter grades to each student for each subject in 2004 and 2005 by calculating the percentile category they are in, and then adding up a total score.

I use the panel of 9th grade MSGE performance to estimate the impact of the policy on the treated middle schools. The number of schools and students and characteristics of all schools and policy schools are presented in [Table 1.1](#). Earlier years have more missing data than later years. Altogether, the eleven policy schools have around 3,000 students per cohort, which is about 14% of all students in the city. Comparing the last two columns across both panels, we can see that policy schools have lower 9th grade scores and lower elite high school attendance rate than other schools across all years.

To evaluate the change in quality of composition, I merge 6th grade baseline records in one district with lottery records for 2007-2010 and 2008-2011 cohorts. The matching rate is higher than 90%.<sup>13</sup> In a separate merge, I match the lottery records in 2005 through 2007 with corresponding MSGE records in 2008 through 2010.<sup>14</sup> The matching rate is higher than 70% across these cohorts. More details on these merges and data set construction can be found in [Appendix](#) .

---

<sup>13</sup>Non-perfect matching rate might due to changing names, typing errors in data, transferring, moving out of the city, etc..

<sup>14</sup>The reason why I do not include 2008-2011 cohort in the lottery analysis is because in 2008, there is a change in the lottery school choice mechanism. Private schools were included in the choice set. Since then, there were two stages of lotteries, first for private schools and then for public schools. One needs to make four choices, one public schools, one private schools, whether to go on to public school lottery if win the private lottery, and whether to go on if lose.

Table 1.1: Administrative Panel Data Description

Entire Sample										
year	# schools	# students	% no missing score	avrg # stu per school	% female	9th grade score	% elite high school			
2004	94	24,007	90.14%	254.39	51.12%	18.70	25.38%			
2005	61	21,062	80.83%	344.28	47.15%	18.65	25.21%			
2006	61	16,014	92.18%	261.52	47.37%	15.90	26.41%			
2007	65	16,120	88.61%	247.00	47.84%	17.93	26.74%			
2008	72	16,967	100.00%	234.65	47.21%	18.28	29.85%			
2009	72	24,763	83.38%	342.93	47.13%	18.14	30.21%			
2010	73	25,580	87.05%	349.41	46.78%	18.27	30.23%			
2011	69	25,296	93.22%	365.61	48.44%	18.62	24.77%			
Policy Schools										
year	# schools	# students	% no missing score	avrg # stu per school	% female	9th grade score	% elite high school			
2004	11	2,989	69.42%	270.73	51.56%	17.51	12.10%			
2005	7	1,862	87.49%	265.00	48.91%	16.35	5.98%			
2006	9	2,109	95.26%	233.33	45.63%	13.81	9.06%			
2007	9	1,964	92.31%	217.22	47.17%	16.03	18.77%			
2008	11	2,023	100.00%	182.91	48.44%	16.26	22.70%			
2009	10	3,100	71.77%	309.00	45.48%	17.25	19.85%			
2010	10	3,025	76.10%	301.50	46.35%	17.40	23.01%			
2011	10	2,398	91.03%	238.80	49.17%	17.47	18.99%			

Note: The table shows the number of schools and students I observe in the administrative data set. There are 11 policy schools in total. Elite high school attendance is available for year 2009 to 2011, and I impute it for the other years by computing the percentage of students with a score higher than the seventieth percentile in the 9th grade graduation exam.

Table 1.2: Individual Level Data: Summary Statistics 2007

Sample		(1) All	(2) policy schools	(3) pre-admitted	(4) oversubscribed lottery	(5) lottery & policy
Academic performance						
	9th grade score	22.73	21.68	24.71	21.60	21.55
	normalized 9th grade score	0.758	0.723	0.824	0.720	0.718
	% passing grade	0.795	0.763	0.876	0.766	0.767
	% academic high school	0.782	0.679	0.935	0.734	0.706
	% elite high school	0.332	0.299	0.529	0.201	0.249
	non-academic evaluation	19.13	18.82	19.55	18.94	18.97
	imputed 9th grade score	22.78	21.93	24.75	21.65	21.69
Family background						
	% female	0.476	0.463	0.481	0.469	0.477
	% with city hukou	0.713	0.650	0.832	0.655	0.617
	% missing hukou status	0.0642	0.0681	0.0478	0.0664	0.0457
	father political	0.239	0.0924	0.454	0.131	0.103
	% missing father's political	0.332	0.274	0.346	0.272	0.207
	mother political	0.119	0.0209	0.256	0.0469	0.0266
	% missing mother's political	0.360	0.294	0.388	0.284	0.216
Middle school admission						
	pre-admission	0.388	0.00345	1	0	0
	over-subscribed lottery	0.286	0.497	0	1	1
	policy school	0.138	1	0.00123	0.240	1
Observations		14,699	2,027	5,709	4,199	1,007

Note: Column 2 describes policy school students, column 3 describes students who were pre-admitted, column 4 describes students who chose an over-subscribed school and assigned by lottery; column 5 describes students who were assigned by lottery to a policy school. % academic high school indicates the percentage of 9th grade graduates attending an academic high school; some other graduates attend vocational schools or stop going to schools. Non-academic evaluation consists of teacher and self-rated measures of four abilities, including civics, learning ability, atheistic ability, and practical ability. Imputed 9th grade score is constructed by assigning the highest grade of their cohort to the missing grade of direct admitted students who did not take the exam. Having city hukou means that a student is born in a city and enjoys the public goods of that city; it is often used as a measure of socioeconomic background. Father and mother political is a dummy that equals one if the parent is affiliated with any party; parental party affiliation is an indicator of better family background.

## 1.4 DOES THE TOP TEN-PERCENT QUOTA POLICY WORK?

### 1.4.1 Comparisons of Trends in Performance

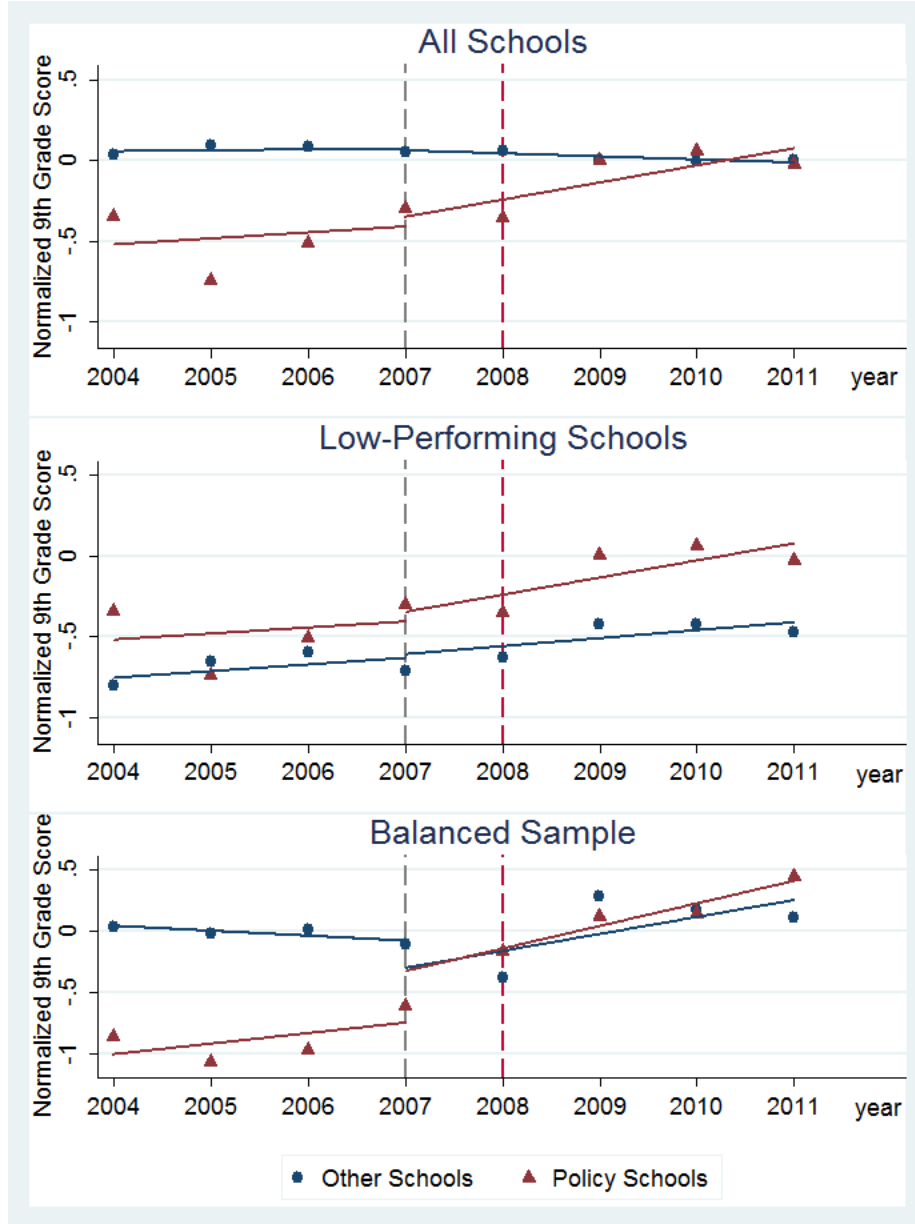
To look at the overall trend in the school performances, I plot the MSGE average scores of the treatment group (policy schools) and that of the control group across years. The exercise is done on three samples: all schools, low-performing schools and a balanced sample. For each sample, the numbers of schools in the treatment and control groups are (11, 93), (11, 37) and (2,20). The entire panel data set has 11 policy schools and 93 other schools.<sup>15</sup> Low-performing schools are defined as schools with average 9th grade score below average in 2004.<sup>16</sup> The balanced sample includes 2 policy schools and 20 other schools, which have MSGE scores records for 2004 through 2011.

---

<sup>15</sup>The sum is larger than the number of schools in any year, because some schools were shut down or merged into other schools and some schools were opened in later years.

<sup>16</sup>Five policy schools did not have data on MSGE scores in 2004, and I assign them to the low-performing schools sample. The fact that they are low-performing is verified by comparing the performance of these five schools with other six policy schools.

Figure 1.2: Normalized 9th Grade Scores: Policy Schools and Other Schools



Notes: Two vertical lines in each figure indicate the timing of the policy change, 2007 and 2008. Three figures differ in the sample they use to plot the scatter and linear fitted lines. The top figure compares the average academic performance of policy schools with all other schools from 2004 to 2011. The middle figure takes low-performing schools with average 9th grade score below medium in 2004 and plots their performances across different years. The bottom figure only uses schools with complete data for each year.

Figure 1.2 shows three sets of comparisons of normalized 9th grade scores for policy schools and the control schools, before and after the introduction of the policy. Each includes scatter plots of group averages and linear fitted lines. Two vertical lines in each figure indicate the timing of the policy change. Three figures differ in the sample they use to plot the scatter and linear fitted lines as defined in the previous paragraph. The top figure uses the entire sample. Before the policy, the performance gap between policy schools and other schools is around 0.5 standard deviations; after the policy, the performance gap was gradually closed.

The middle figure uses only the low-performing schools to make the comparison. Before the introduction of top ten-percent quota policy in 2007, policy schools' average academic performances were slightly higher than other low-performing schools; both sets of schools performed below average with normalized standardized scores at around -0.6 -0.5 and had an improving trend. After that, the policy schools improved more rapidly and increased their average scores to above average, while the other low-performing schools improved slightly but still had normalized average performance at around -0.5.

The bottom figure uses the balanced sample. Since previous years have more occurrences of missing data, we may worry if data is missing for policy schools when they coincidentally performed well or bad in that pre-policy year, which would bias the treatment effect upward or downward. The balanced sample only includes 22 schools and may not give us the accurate estimates of the impact. Therefore, using this sample serves as a robustness check to make sure the main results are not due to accidental biases from an unbalanced panel. For the schools with no missing data, before 2007 policy schools had a slight improving trend; after 2007 they made a parallel movement upward by around 0.5 standard deviations and ended at around average score.

### 1.4.2 Test for Selection in Treatment Status

Not all low-performing schools received the treatment of the top ten-percent quota policy.<sup>17</sup> This is a result of negotiation between the education bureau and the elite high schools. Elite high schools were only willing to set aside a limited number of seats for unconditional acceptance of top Ten-Percent students from low-performing schools.

To verify that the policy schools were not chosen based on observed characteristics, I run a set of probit regressions to test whether pre-policy characteristics can predict policy treatment status.

$$D[policy]_j = \alpha_j + \alpha_1 X_j + \epsilon_{jt} \quad (1.1)$$

$D[policy]_j$  is a dummy variable for policy treatment status. It is separately defined by whether a school is treated by the Top Ten-Percent Policy in 2007, in 2008 or in either year.  $X_j$  represents a set of pre-policy characteristics, including normalized MSGE average performances, normalized school rankings, female percentages, and numbers of students. Marginal effects are reported in [Table 1.3](#). First, I run the analysis for all schools and report the results in column (1)-(3). None of the coefficients are statistically different from zero, which indicates that observed school characteristics fail to predict the assignment of the top ten-percent quota policy. To obtain a more concise decision faced by education bureau and elite high schools, I then restrict the sample to only low-performing schools. Column (4)-(6) reports the results. All observed school characteristics still fail to predict treatment status.

This exercise tests for selection on the observed school characteristics and fails to find any. If the education bureau assigned quotas to the policy schools in 2007 because these schools would grow faster after 2007 for unobserved reasons, regardless of the top ten-percent quota policy, the results would be biased. However, given the objective of equalizing school performance, the education bureau would not have chosen these schools if they were

---

<sup>17</sup>Recall from the policy background section that six schools have been treated since 2007 and five other schools since 2008.

Table 1.3: Test for Treatment Status Selection

	(1)	(2)	(3)	(4)	(5)	(6)
Sample	All Schools			Low Performing Schools		
Dependent Variable	2007 Policy	2008 Policy	Any Policy	2007 Policy	2008 Policy	Any Policy
normalized MSGE score	-0.0334 (0.0447)	-0.0261 (0.0411)	-0.0559 (0.0575)	-0.0191 (0.0799)	-0.000592 (0.0591)	-0.0196 (0.0956)
number of students	1.62e-05 (0.000103)	3.21e-05 (7.97e-05)	5.60e-05 (0.000121)	3.90e-06 (0.000179)	-2.31e-06 (0.000138)	-1.72e-05 (0.000217)
percent female	0.343 (0.253)	0.0190 (0.250)	0.374 (0.339)	0.529 (0.432)	-0.0280 (0.349)	0.532 (0.524)
normalized ranking	-0.169 (0.110)	-0.00633 (0.0975)	-0.170 (0.135)	-0.0726 (0.214)	0.0506 (0.160)	-0.0261 (0.257)
Obs (School by year)	208	208	208	100	100	100
Pseudo R2	0.157	0.0144	0.0865	0.0303	0.00397	0.0132

This table reports probit regression results, testing for whether pre-policy school characteristics can predict policy treatment status. Dependent variables are indicators for whether a school is treated by the Top Ten-Percent Policy in 2007, in 2008 or in either year. The sample used for column (1)-(3) is all observations in the year 2004, 2005 and 2006; the sample used for column (4)-(6) is low-performing schools in those three years. Marginal effects and standard errors in parentheses are reported, \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . All these characteristics fail to predict treatment status.



already improving by themselves. Instead, they would have chosen other schools that needed help. Therefore, it is implausible that the policy schools were assigned treatment based on observed or unobserved characteristics.

### 1.4.3 Difference-in-Differences Estimation

In general, the pre-trends of treatment and control groups are parallel in all three plots in [Figure 1.2](#), which verifies the parallel trend assumption needed for a difference-in-differences estimation. In addition, pre-policy school characteristics fail to predict treatment status of a school, which indicates that policy schools were not chosen based on these observable characteristics and can be deemed as an exogenous shock. To obtain a magnitude of the impact, I use the following difference-in-differences specification:

$$Y_{it} = \alpha_i + \delta_t + \beta D(policy)_i * D(post)_t + \epsilon_{it} \quad (1.2)$$

$Y_{it}$  stands for school  $i$ 's performance measures, including normalized average 9th grade score, normalized ranking and percentage of students attending elite high schools in year  $t$ . Normalized average 9th grade score has a mean of 0 and a standard deviation of 1 for every year. Normalized ranking is constructed by dividing the school ranking (the higher the better) with total number of schools that year. I add school fixed effects  $\alpha_i$  and year fixed effects  $\delta_t$  to control for differences in exams and cohorts across years and fixed differences between schools. The coefficient of interest is  $\beta$ , which is the impact of the policy on the outcome variables  $Y_{it}$ . The dummy indicator  $D(policy)_i * D(post)_t$  takes value of one when school  $i$  had the quota in year  $t$ . For the six policy schools assigned the quota in 2007, this dummy is one for year 2007 onward; for the five policy schools assigned the quota in 2008, it is one for year 2008 onward; for all other schools, it is zero for all years. An alternative specification is replacing year fixed effect dummies with a linear time trend.

For similar reasons explained in the previous subsection, I do the analysis for three

Table 1.4: Difference-in-differences: Treatment Effect on School Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)
Depend.Variable	Normalized 9th grade score	Normalized school ranking			% elite high school	
Panel A. All Schools						
postXpolicyany	0.357*** (0.120)	0.296** (0.120)	0.116*** (0.0389)	0.0965** (0.0387)	0.0631*** (0.0200)	0.0695*** (0.0205)
Observations	556	556	556	556	556	556
R-squared	0.070	0.027	0.064	0.028	0.154	0.066
Number of schools	104	104	104	104	104	104
Panel B. Low-performing Schools						
postXpolicyany	0.309** (0.124)	0.250** (0.121)	0.0666 (0.0434)	0.0517 (0.0417)	0.0497*** (0.0170)	0.0601*** (0.0172)
Observations	275	275	275	275	275	275
R-squared	0.102	0.062	0.108	0.089	0.277	0.184
Number of schools	48	48	48	48	48	48
Panel C. Balanced sample						
postXpolicyany	0.939*** (0.249)	0.875*** (0.261)	0.327*** (0.0815)	0.305*** (0.0846)	0.0794* (0.0425)	0.0847* (0.0451)
Observations	176	176	176	176	176	176
R-squared	0.230	0.106	0.213	0.105	0.246	0.101
Number of schools	22	22	22	22	22	22
School FE	Y	Y	Y	Y	Y	Y
Year FE	Y	N	Y	N	Y	N
Time Trend	N	Y	N	Y	N	Y

Notes: Standard errors in parentheses. Significance level indicated by \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . The time span of the school panel data ranges from 2004 to 2011. “postXpolicy” equals one for policy schools during the years after the initial effective year. Each column has dependent variable listed on top row and control variables indicated at the bottom three rows. Panel B takes school with lower than average score in 2004; Panel C takes schools with observed performance in all eight years.

samples. Results are shown in [Table 1.4](#). Panel A, B, and C look at all schools, low-performing schools and a balanced sample, respectively. Almost all coefficients are positive and significant. Estimates using the balanced panel are the largest and those using only low-performing schools are the smallest. Compared with all other schools, policy schools experienced 0.3 standard deviation improvement in 9th grade school average scores. Their school ranking also rose by 11 to 15 percentile, which would mean moving up seven to eight spots among a ranking of around 80 schools. Their elite high school attendance increased significantly for about 6%. Results do not vary much with year fixed effects or a linear time trend.

While the interacted dummy variable in [Equation 1.2](#) captures the average treatment effect on the policy schools, it assumes a one time shift in the performances instead of gradual changes. The following alternative specification uses years treated instead of binary indicator and looks at the incremental effect of the policy.

$$Y_{it} = \alpha_i + \delta_t + \beta T_{it} + \epsilon_{it} \tag{1.3}$$

$T_{it}$  indicates years treated. It takes value of  $\max\{0, t - 2006\}$  for 2007 policy schools,  $\max\{0, t - 2007\}$  for 2008 policy schools and zero for all other schools. Similarly, I conduct the analysis on all three samples.

[Table 1.5](#) shows that almost all coefficients are positive and significant. Similar with patterns in [Table 1.4](#), estimates using balanced panel are the largest and those using only low-performing schools are the smallest. Compared with all other schools, policy schools improved by 0.06 standard deviation every year. Their school ranking also rose by 2 percentile each year, which would mean moving up in the ranking by one every year. The increase in elite high school attendance is around 1.6% every year.

Table 1.5: Incremental Treatment Effect on School Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)
Depend.Variable	Normalized 9th grade score	Normalized school ranking			% elite high school	
Panel A. All Schools						
treat year	0.0646** (0.0264)	0.0663** (0.0267)	0.0232*** (0.00854)	0.0235*** (0.00860)	0.0161*** (0.00438)	0.0168*** (0.00456)
Observations	556	556	556	556	556	556
R-squared	0.064	0.027	0.061	0.030	0.161	0.071
Number of schools	104	104	104	104	104	104
Panel B. Low-performing Schools						
treat year	0.0567** (0.0277)	0.0555** (0.0276)	0.0123 (0.00965)	0.0118 (0.00955)	0.0145*** (0.00372)	0.0149*** (0.00392)
Observations	275	275	275	275	275	275
R-squared	0.094	0.061	0.105	0.089	0.298	0.191
Number of schools	48	48	48	48	48	48
Panel C. Balanced sample						
treat year	0.201*** (0.0537)	0.201*** (0.0565)	0.0741*** (0.0175)	0.0741*** (0.0182)	0.0199** (0.00914)	0.0199** (0.00980)
Observations	176	176	176	176	176	176
R-squared	0.229	0.114	0.221	0.124	0.252	0.104
Number of schools	22	22	22	22	22	22
School FE	Y	Y	Y	Y	Y	Y
Year FE	Y	N	Y	N	Y	N
Time Trend	N	Y	N	Y	N	Y

Notes: Standard errors in parentheses. Significance level indicated by \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . The time span of the school panel data ranges from 2004 to 2011. “Treat year” equals to zero for all school year combination where there were no quota policy and equals number of years since quota policy is effective for policy schools. Each column has dependent variable listed on top row and control variables indicated at the bottom three rows. Panel B takes school with lower than average score in 2004; Panel C takes schools with observed performance in all eight years.

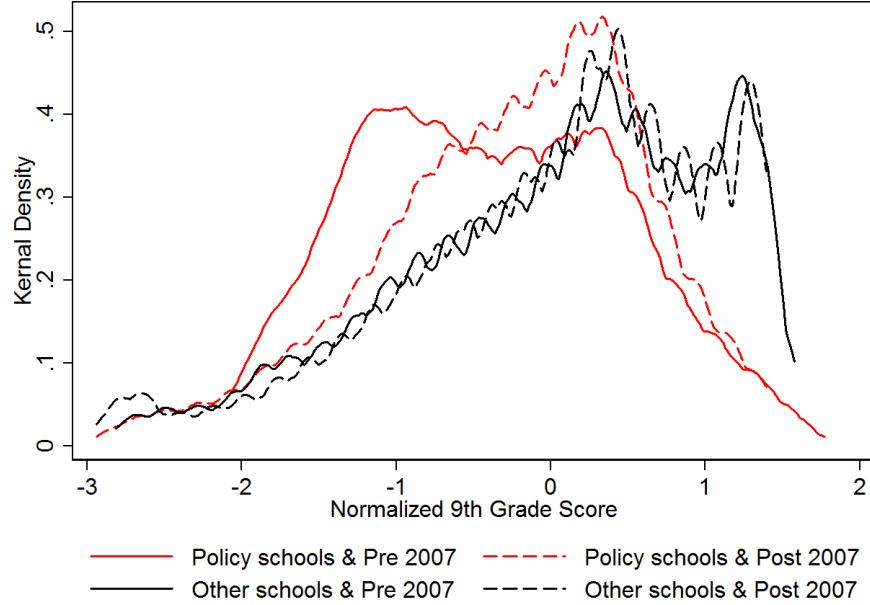
#### 1.4.4 Placebo Test

To verify the treatment effects only exist for the policy schools instead of for any low-performing schools, I run a falsification test of [Equation 1.2](#) on the panel data without policy schools, and assign a placebo treatment status to other low-performing schools. Coefficients of the interaction dummies are reported in the Appendix Table A4. Placebo treatment effects are statistically insignificant in all except one specification: outcome variable being normalized ranking and controls including school and year fixed effects. This effect goes away after including a linear time trend, which indicates that the only statistically significant placebo treatment effect is because of variable construction. Normalized school ranking is constructed as ranking divided by total number of schools that year, and total number of schools decreases from 2004 to 2011 overall. Looking back at actual treatment effect estimates in [Table 1.4](#), including the linear time trend did not change the results. Therefore, the falsification test suggests that treatment effects only apply to the policy schools.

#### 1.4.5 Change in Distributions of Academic Performances

I plot the 9th grade score distributions for policy schools and non-policy schools before and after 2007 in [Figure 1.3](#). Solid lines are for before 2007 and dotted lines are for after 2007; red lines are for policy schools and black ones are for non-policy schools. To see the change in distributions of policy schools' normalized 9th grade scores, I compare the dotted and solid red lines. There is a clear rightward shift of policy schools' score distribution, and the Ksmirnov test rejects that two distributions are the same. This means that policy schools improved their performance. A smaller proportion of students get a normalized score one standard deviation below mean and more get a score around the average. On the other hand, the difference between distributions of non-policy schools before and after 2007 is not clear by looking at the two black lines in the graph. To understand the change

Figure 1.3: General Pattern of Changes in Score Distributions



Notes: The figure plots individual level normalized 9th grade scores for policy schools and non-policy schools before and after 2007.

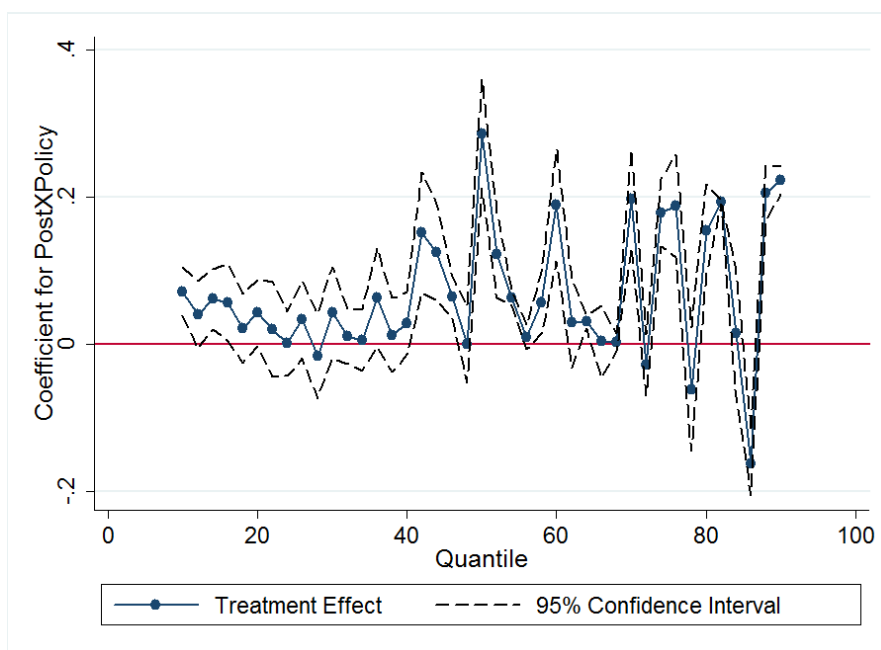
in distributions, I run quantile regressions with the following specification:

$$Y_{it} = \alpha_i^p + \delta_t^p + \beta^p D(policy)_i * D(post)_t + \epsilon_{it}^p \quad (1.4)$$

where  $0 < p < 1$  indicates the proportion of the population having scores below the  $p$ th percentile. Estimates for different quantiles are plotted in [Figure 1.4](#), with 95% confidence intervals around them. The estimates are very volatile, especially for the top percentiles. This may be due to the discrete nature of translated letter grades. For the lower end of the distribution, there is not much change in the performance; for the middle part, we see positive changes for some quantiles and insignificant changes for others; for the top end, estimates are very volatile. It suggests that students at the middle experienced significant

improvement in their 9th grade performance, while students at the bottom did not. The story is less clear for the top end of the distribution. The quantile regression results only show a noisy picture of changes along the distribution and do not offer any insight on selection and change in school quality.

Figure 1.4: Quantile Regression Estimates



Notes: The figure plots estimates of the interaction term `postXpolicy` in quantile regressions from 10 percentile to 90 percentile with a 2 percentile increment.

## 1.5 UNDERLYING MECHANISMS

### 1.5.1 A Conceptual Framework

I documented in the previous section that the gap between the policy impacted low-performing middle schools and the others has been shrinking after the policy. The top

ten-percent quota policy might work through three interacting channels. First, it increases the chance to be admitted into a top high school for students attending the policy schools. Therefore it encourages 6th grade elementary school graduates, especially the high performing ones, to choose the policy impacted schools. I refer to this as a *composition effect*. Secondly, if the first channel exists, having better peers may bring a positive effect to students attending the policy schools (Ding and Lehrer (2007)). I refer to this as a *peer effect*. Thirdly, competition for the top ten-percent within these schools may also help the schools improve their average performance. I refer to this as a *tournament effect*. Top-performing students in these schools could be affected by both peer effects and a tournament effect; while low-performing students are most likely affected by peer effects but not the tournament effect, since they know that placing at the top ten-percent is unlikely.

To understand what drives the impact, I modify the theoretical framework from Cullen et al. (2013) to help illustrate the mechanisms and motivate the empirical tests for changes in composition. For simplicity, I abstract away from heterogeneous neighborhood characteristics, transportation cost and tuition and assume they are identical. This assumption is plausible because of the following reasons: the public middle schools one can choose from are nearby and in the same district; public transportation is cheap and convenient; this city has rather low crime across all districts; there is little ethnic variation across neighborhoods; public middle school tuition is regulated (around USD50 per semester).

The decision of school choice by parents is driven by the expected impact schools will have on their children’s future prospects. In China, graduating from an elite college gives a high return rate (Li et al. (2012)). Therefore, I set the objective of school choice to maximize college entrance exam performance,  $Y_{ij}$ . It depends on student’s innate ability,  $a_i$ , learning progress before entering high school,  $y_{ij}$ , and probability of attending an elite high school,  $p_{ij}$ . I assume learning progress in middle school increases with one’s ability  $a_i$ , peer quality  $q_j$ , and school characteristics and learning atmosphere  $\gamma_j$ . Probability of attending an elite high school is included because elite high schools have better school



quality and peers, and a significantly higher percentages of students attending an elite college.

$$\max_j Y_{ij} = Y_{ij}(a_i, y_{ij}(a_i, q_j, \gamma_j), p_{ij}(y_{ij})) \text{ for } j = 1, 2 \dots n \quad (1.5)$$

Without loss of generality, assume that families face two public middle school choices,  $j = 1, 2$ , and school 1 has a better peer quality  $q_1 > q_2$ , and  $y_{i1} < y_{i2}$ . Suppose that before the policy, probability of getting into an elite high school is higher in school 1 for student  $i$ ,  $p_{i1} < p_{i2}$ . The policy gives school 2 elite high school admission for top ten-percent and changes  $p_{i2}(y_{i2}, \kappa_{i2})$ , where  $\kappa_{i2}$  means the probability of placing at top ten-percent in school 2. Now, for a small set of students, the policy changes the comparison between probability of attending an elite high school in school 1 and school 2,  $p_{i1}(y_{i1}) < p_{i2}(y_{i2}, \kappa_{i2})$ . These students switch to school 2 when the benefit from higher probability of attending an elite high school is larger than the cost of having a lower peer quality,  $\frac{\partial Y_{ij}}{\partial p_{ij}} > \frac{\partial Y_{ij}}{\partial y_{ij}} \frac{\partial y_{ij}}{\partial q_j}$ .

On the other hand, for some top performing students, the probability of getting into an elite high school is  $p_{ij}(y_{ij}) = 1$ , and they have no incentive to bear cost in downgrading to school 2 with lower peer quality.

To sum up, the introduction of top ten-percent quota policy changes  $p_{ij}$  for policy schools and alter some students' school choices. This affects policy schools in two dimensions: change in peer composition  $q_j$  and change in value-added  $y_{ij}$  caused by change in peer composition  $q_j$  and competitive learning atmosphere  $\gamma_j$ .

### 1.5.2 Using Lottery Records to Tease Out Mechanisms

Empirically, the lottery middle school assignment provides the opportunity to separately look at changes in composition and in value-added. Recall that each elementary school is assigned a certain number of seats to around three nearby middle schools. Each 6th grader chooses only one middle school. When a middle school gets oversubscribed in that elementary school, a lottery randomly determines school assignment.

Policy schools rarely gets oversubscribed in any elementary schools. Therefore, there are only two types of 6th graders who attend policy middle schools: those who voluntarily enroll in policy schools, and those who choose some other school but lose the lottery and get randomly assigned to a policy school. The composition effect mainly captures changes in the ability of the first type, “voluntary enrollees”. I analyze the baseline scores of sixth graders who voluntarily chose a policy school and test if there is evidence of strategic switching after the policy by high-performing students.

For the second type, “lottery losers”, they did not strategically change their school choice after the policy, yet they still may benefit from attending a policy school compared to previous cohorts, because now they have better peers. I exploit the random lottery assignment to evaluate differences in value-added between the policy schools and the chosen schools, before and after the policy, to see how the value-added gap was changed by the policy-induced peer composition change. High-performing lottery losers may benefit more from good peers and a tournament incentive than lower-performing lottery losers, therefore I conducted instrumental quantile treatment effect analysis to detect differential treatment effects along the distribution.

### **1.5.3 Change in Composition**

The top ten-percent quota policy increases the expected return of attending the policy schools, changes the trade-off of school choice, and therefore incentivizes some sixth graders to strategically choose the policy schools. As illustrated in the conceptual framework, switching to a policy school involves trade-offs. A student may benefit from the expectation that he/she will be in the top ten-percent of the graduating class, but at the same time may suffer from a lower peer quality. Therefore, it is unclear whether students would respond to the policy by changing their school choice.

The composition effect has two dimensions, quantity and quality. Higher percentages of sixth graders may voluntarily choose the policy; among those sixth graders who choose

policy schools, the average baseline performance may be higher than before. First, to test the quantity dimension of demand change, I use four years of lottery choice records from 2005 to 2008 and compute the percentages of students choosing the policy-impacted schools for each elementary school. I run the following regression to check if there was a significant increase in the percentages of students choosing the policy schools.

$$Perc(S_{jt}) = \theta_j + \beta D(post)_t + \epsilon_{jt} \quad (1.6)$$

$Perc(S_{jt})$  is the percentage of 6th graders choosing policy school in elementary school  $j$  in year  $t$ ;  $\theta_j$  stands for elementary school fixed effect;  $D(post)_t$  is a dummy for post policy years.  $\beta$  is the coefficient of interest, which indicates how many more students, in percentage of their elementary schools, chose policy schools after the change. [Table 1.6](#) shows that the percentages of students who chose policy schools did not significantly change after the policy announcement. Therefore, there was not a significant increase in the quantity of students voluntarily enrolled in policy schools.

To test the change in composition quality, I analyze the incoming students' sixth grade scores. Since I only have the sixth grade scores for one district in 2007-2010 and 2008-2011 cohorts, I compare the change in scores of elementary students choosing the policy school announced in 2008 for that district. To verify the generalizability of results from this district, I compare the differences in school characteristics and policy impacts of this district and other districts. Results show that the comparison between policy and non-policy schools and policy impacts are not statistically different from those of other districts. The conditional logistic regression with elementary school fixed effects is specified as below:

$$Pr(S_i = 1 | \mathbf{x}_i, D(post)_t) = F(\alpha_i + \mathbf{x}_i\beta_1 + \beta_2 D(post)_t + D(post)_t * \mathbf{x}_i\beta_3) \quad (1.7)$$

where  $F$  is the cumulative logistic distribution,  $F(z) = \frac{\exp(z)}{1+\exp(z)}$ .  $S_i$  equals to 1 if student  $i$  chose a 2008 policy school in year  $t$ ;  $\alpha_i$  stands for elementary school fixed effect;  $D(post)_{it}$  is a dummy for year 2008 and  $\mathbf{x}_{it}$  is a vector of student  $i$ 's 6th grade math score and reading

Table 1.6: Composition effect

## Change in Percentages of Sixth Graders Choosing Policy Schools

Sample	(1) All Elementary schools	(2) All Elementary schools	(3) Elementary schools with Non-Zero Percentage Choosing Policy Schools	(4) Elementary schools with Non-Zero Percentage Choosing Policy Schools
Dependent Variable	% of Students Choosing 2007 Policy Schools	% of Students Choosing Any Policy School	% of Students Choosing 2007 Policy Schools	% of Students Choosing Any Policy School
After 2007	-0.0131 [0.0107]		-0.0103 [0.0203]	
After 2008		-0.00745 [0.0133]		-0.00618 [0.0170]
Constant	0.0838*** [0.00752]	0.144*** [0.00659]	0.316*** [0.0131]	0.326*** [0.00783]
Elementary School FE	Y	Y	Y	Y
Observations	979	979	243	430
R-squared	0.617	0.722	0.931	0.907

Standard errors in brackets, \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

Sample includes 2005-2008 four cohorts of elementary school graduates' school choice. *After 2007* equals to 1 if the year is after 2007 and *After 2008* equals to 1 if the year is after 2008.

Table 1.7: Are elementary students with high ability and high SES more likely to choose a policy school after the policy?

	(1)	(2)	(3)	(4)	(5)
Dependent Variable	choose a policy school				
postXmath	0.049** (0.0174)				
postXreading		0.0217 (.0297)			
postXfather politics			0.0055** (0.00203)		
postXmother politics				0.0095** (0.0045)	
postXfemale					0.017 (0.015)
Observations	4,151	4,151	2,746	2,499	4,570

Each column reports the marginal effect of the interaction term in a fixed effect logit regression controlling for elementary school fixed effects, corresponding to [Table 1.7](#). Robust standard errors were allowed to clustered at the elementary school level. Significance level is indicated by \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . Father and mother political statuses are dummy variables for any parental political affiliation, which are proxies for high socioeconomic status. The table shows positive and statistically significant changes in family background and baseline academic performance for students who voluntarily enroll in policy schools.

Table 1.8: Conditional logit estimates of choosing a policy school: Who Are Switching?

Dependent Variable	Choosing a policy school	
	(1)	(2)
6th grade score of	math	reading+math
post X top1%	-0.381	-0.159
post X top5%	0.877	0.096
post X top10%	1.341**	1.580**
post X top20%	1.130**	0.702
post X top30%	0.794	0.635
post X top40%	1.344**	0.925
post X top50%	0.589	1.611***
Elementary School FE	Y	Y
Observations	4,570	4,570

Robust standard errors in parentheses, clustered at the elementary school level. Significance level:\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

This table reports coefficients from fixed-effects logit regressions. The coefficients can not be interpreted as magnitudes of changes in likelihood of choosing a policy school. Significance level and signs of the coefficients provides information on who are switching. Sample used is merged lottery records with sixth grade scores from one district in 2007 and 2008. In 2008, one middle school was impacted by the policy in this district. The dependent variable equals to one if a sixth grader chose the 2008 policy school. “post” equals to 1 for the year of 2008. “top x%” equals to 1 if a sixth grade’s score is in the top x% in their cohort, all “top x%” categories are mutually exclusive and all regressions include dummies for each category.

(Chinese) score. The coefficient of interest is  $\beta_3$ , which tells us how the ability of incoming students in the policy schools changed after the policy was announced.

Table 1.7 shows the regression results for Equation 1.7. Marginal effects on the interaction terms between post policy and student characteristics are reported. The interaction terms of normalized sixth grade math score and dummies for parental political affiliations are highly significant and positive. It indicates that students who have better math score and better socioeconomic background are more likely to choosing policy schools after the introduction of the ten-percent quota policy.

Although we now know that students with higher baseline math scores switch to choose policy schools, we do not know if it is because of a heavy downgrade by a few top students, or a small downgrade by many medium-ranking students. By analyzing the trade-off faced by these students, the highest-performing students have less incentive to downgrade than the second-tier-ranking students because the highest-performing students are confident about getting elite high school admission, even without the policy guarantee. Therefore, they have no benefit and no incentive to pay the cost of having lower quality peers by switching to a policy school.

To test this hypothesis and understand the characteristics of the students who switch, I use the same regression specification as Equation 1.7 and replace the actual baseline score with the percentile category dummies that are mutually exclusive. For example, if the top 5% takes value of 1, it means that a student has a baseline score between 1% to 5%. Estimates in Table 1.8 confirm the hypothesis that the highest-performing 6th graders were not more likely to switch, while sixth graders with above-average math scores showed statistically significantly switching patterns.

#### 1.5.4 Change in Value-added

In the previous subsection, I show that the top ten-percent quota policy attracts students with better math scores to voluntarily enroll in the treated lower-performing schools. One

of the concerns is that the policy may have improved school average performance only by redistributing students, without changing the school quality at all. After the policy, there are three types of students in the policy schools: strategic switchers, students who would have chosen the policy schools anyway and students who choose an over-subscribed school but lose the lottery and get randomly assigned to a policy school. The previous subsection analyzes school choices by the first two types of students; this subsection takes the last type of students, lottery losers, and compare them with the lottery winners to estimate value-added gaps between policy schools and over-subscribed schools.

Change in value-added can come from several channels. First, policy schools may bring higher value-added to the cohorts entering middle school after the policy than previous cohorts, because of better peers attracted by the policy. In addition, competition to place at the top ten-percent of the graduating class encourage students to exert more effort. Especially since policy schools determine ranking by three-year accumulated performance, students need to work consistently throughout the three years. High-performing students may benefit more from tournament incentive than lower-performing ones.

Different from school performance, school quality is usually difficult to measure because of endogenous selection. Better students often sort into schools with better reputation, which makes it hard to disentangle whether the higher performance in these schools comes from incoming students' ability or school quality. The random lottery assignment allows me to use it as an instrument to evaluate differences in value added with the Local Average Treatment Effect (LATE) model by [Imbens and Angrist \(1994\)](#). These LATE estimates provide measurements of the school quality gap before and after the policy introduction and therefore enable us to see changes in value-added.

The instrumental analysis uses the sample of students who chose an over-subscribed school and randomly assigned to their choice schools or an under-subscribed policy school. Let  $Y_i(1)$  be student  $i$ 's potential test score if she attends a policy school, and let  $Y_i(0)$  be her test score if she attends her choice school.  $D_i$  indicates the "treatment", policy



school attendance, and  $Z_i$  is an indicator for lottery outcome. Let  $D_i(1)$  and  $D_i(0)$  denote potential treatment status as a function of  $Z_i$ . The following assumptions are needed for LATE framework:

1. Independence and Exclusion Restriction:  $(Y_i(1); Y_i(0); D_i(1); D_i(0))$  is independent of  $Z_i$ .
2. Nontrivial First Stage:  $Pr(Z_i) = E[D_i|Z_i]$ .
3. Monotonicity:  $D_i(1) > D_i(0)$  for all  $i$ .

The first assumption requires that lotteries are random and do not affect test scores through any channel but policy school attendance. The second assumption requires that lottery losers are more likely to attend policy schools on average. Monotonicity assumption requires that winning the lottery does not encourage any student to attend a policy school instead of the choice school. All three assumptions are satisfied in this study's sample.

First, to verify the random lottery school assignment, I use a probit model to regress students' pre-lottery characteristics on their lottery outcomes. If the lotteries are random, pre-lottery characteristics should not be able to predict the lottery outcome. I include a group of dummy variables to control for the lottery choice and the elementary school attended, since lotteries happen at the elementary school level. The regression equation is

$$Z_{ic} = \alpha_c + \alpha_1 X_i + \epsilon_{ic} \quad (1.8)$$

$Z_{ic}$  equals 1 if the student  $i$  lost the lottery.  $X_i$  represents pre-lottery characteristics including gender, city residency (hukou), parental political status, Chinese and math scores in elementary school graduation exam. Parental political status and elementary school graduation exam scores are only available for 2007, not for 2005 or 2006.

[Table 1.9](#) reports the marginal effects of  $X_i$  and verifies the lottery randomness for 2005-2008 and 2007-2010 cohorts. Pre-lottery characteristics cannot predict lottery outcomes for these two cohorts. Conditional on taking the lottery, winning the lottery is an exogenous event that sends students who made the same lottery choice into different middle schools in

these two years. This gives us a device to peel away the endogenous school choice problem and compare the value added of the policy impacted ones with the not impacted ones. Results don't change if I put more explanatory variables, for example sixth grade scores and family background, in 2007-2010 cohort. I do not include 2006-2009 cohort in the 2SLS and instrumental quantile treatment effect analysis due to missing data.

This finding is of central importance for this paper, because one of the concerns for the policy is that it may have improved school average performance only by redistributing students and may have not improved the school quality of these low-performing schools at all. The analysis here shows that before the policy, among the students who chose an over-subscribed school, lottery losers who were randomly assigned to a low-performing policy school performed worse than their elementary school classmates who won the lottery and assigned to an over-subscribed school in the middle school graduation exams. After the policy, their average outcomes were about the same. To verify the predictive power of losing a lottery on attending a policy school, I run the first stage probit regression, controlling for lottery choice group fixed effects and available student characteristics. [Table 1.10](#) reports the marginal effects and the Pseudo R squared, which indicates that using lottery outcomes to instrument for policy school attendance is nontrivial.

$$D_i = \kappa_c + \alpha_1 Z_i + X_i \alpha_2 + \mu_{ic} \quad (1.9)$$

After verifying the assumptions needed for the LATE framework, students can be divided into three types: always takers, who attend regardless of the lottery outcomes ( $D_i(1) = D_i(0) = 1$ ), never takers, who never attend policy schools ( $D_i(1) = D_i(0) = 0$ ), and compliers, who are induced to attend by receiving offers ( $D_i(1) > D_i(0)$ ). The instrumental variables methods can consistently estimate LATE, the average treatment effect for compliers ([Imbens and Angrist, 1994](#)):

$$\frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]} = E[Y_i(1) - Y_i(0)|D_i(1) > D_i(0)] \quad (1.10)$$

I use a two-stage least squares (2SLS) method to estimate LATE. The regression specifi-

Table 1.9: Lottery Randomness Verification

Dependent Variable: Winning a Lottery			
Year	2005		2007
female	0.0254 (0.0180)	-0.00228 (0.0151)	0.0115 (0.0175)
hukou		-0.0208 (0.0176)	0.0157 (0.0204)
father political status			0.0381 (0.0296)
mother political status			0.0606 (0.0466)
Lottery fixed effects	Y	Y	Y
Observations	2,558	2,747	2,747

Robust standard errors in parentheses clustered at the lottery level, \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ . Lottery fixed effects are interacted dummy variables with elementary schools and middle school choice. “Hukou” means whether the student has city residency or not. “Father (mother) political status” indicates whether the student’s parent has a party affiliation or not.

cation is

$$Y_i = \theta_c + \beta_1 D_i + X_i \beta_2 + \epsilon_{ic} \quad (1.11)$$

where  $Y_i$  is the normalized middle school graduation exam score for student  $i$ ,  $D_i$  is a dummy variable indicating policy school attendance, and  $X_i$  is a set of elementary school and lottery choice indicators and student characteristics. The first stage is specified in [Equation 1.9](#).

[Table 1.11](#) reports the results of the 2SLS. The observations are less than the total number of students participating in school lottery assignment, because only lotteries involving a policy school have variations in first stage outcome, i.e. policy school attendance. The comparison between 2SLS estimates for two cohorts shows that the gap between value-added by policy schools and oversubscribed schools were closed by the top ten-percent quota policy.

As discussed at the beginning of this subsection, there are several channels that the policy could have helped close the value-added gap. Peer effects could help improve value-added to all students; a tournament effect could increase value-added to high-performing students who have a chance to compete for the top-ten percent quota. Therefore, if we observe an value-added improvement to the low-performing students, that would be evidence that peer effects were at work.

The LATE masks the heterogeneous treatment effects across students with different academic performance. To see whether peer effects were at work, it is important to estimate the treatment effect across the distribution. Here I use instrumental Quantile Treatment Effect analysis ([Abadie et al. \(2002\)](#)) to analyze the gaps in distributions of value-added between policy schools and chosen middle schools. Similar with 2SLS, this exercise is carried out for two cohorts, one before the policy and one after the policy.

The instrumental QTE is “an Abadie-type weighting estimator of the causal effect of treatment on quantiles for compliers” ([Angrist and Pischke, 2008](#)). The relationship between the QTE estimator and quantile regression is analogous to that between 2SLS

Table 1.10: First Stage: Use losing a lottery to instrument for policy school attendance

Dependent Variable: attending a policy school		
Cohort	2005-2008	2007-2010
losing a lottery	0.244*** (0.0237)	0.393*** (0.0116)
female	0.0245 (0.0246)	0.0001 (0.0160)
hukou		0.0202 (0.0192)
Lottery fixed effects	Y	Y
Observations	1,130	2,066
Pseudo R2	0.249	0.361

Robust standard errors in parentheses clustered at the lottery level, \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. Lottery fixed effects are interacted dummy variables with elementary schools and middle school choice. hukou means whether the student has city residency or not.

Table 1.11: Average Change in School Quality Gap: 2SLS Results

Dependent Variable: Normalized 9th grade score		
	(1)	(2)
Cohort	2005-2008	2007-2010
attending a policy school	-0.302*** (0.0641)	-0.0805 (0.0495)
Obs	1,130	2,066

Standard errors in parentheses, \*\*\* p<0.01. This table reports results of 2SLS regressions on two cohorts, before and after policy respectively. Each coefficient shows the average school value-added gap between policy schools and over-subscribed schools.

and OLS. The set up for QTE estimation is described as following. A scalar outcome variable  $Y$  is students' normalized ninth grade middle school graduation scores.  $D$  is a binary treatment indicator for policy school attendance, and  $Z$  is a binary instrument for losing a lottery.  $X$  stands for a set of dummies for elementary school and school choice, and other student characteristics.

$$Q_{\theta}(Y_i|X_i, D_i, D_{1i} > D_{0i}) = \alpha_{\theta}D_i + X_i'\beta_{\theta} \quad (1.12)$$

where  $Q_{\theta}(Y_i|X_i, D_i, D_{1i} > D_{0i})$  denotes the  $\theta$ -quantile of 9th grade score conditional on control variables  $X_i$  and policy school attendance  $D_i$  for compliers.

The QTE estimation results for Equation 1.12 are reported in Table 1.12 and plotted in Figure 1.5 for 2005-2008 and 2007-2010 cohorts. Looking at the 2005-2008 cohort, we see a larger value-added gap for high-performing students. This is consistent with previous findings in peer effects in Chinese secondary schools by Ding and Lehrer (2007). High-

performing students benefit more from attending over-subscribed schools, which provide better peer quality. For lottery losers who chose another middle school and were randomly assigned to a policy school in 2005, before the ten-percent quota policy, the gap between the distribution of 9th grade graduation scores and the lottery winners' distribution was significantly negative for seven out of nine deciles. After the policy, however, most estimates are insignificantly different from zero, which indicates that policy schools and over-subscribed schools then have similar value-added. For 60th and 90th percentiles, the value-added by policy schools for 2007-2010 cohort were still lower than over-subscribed schools, but less so than the 2005-2008 cohort.

Table 1.12: Change in Distributions of School Quality Gap: IV Quantile Treatment Effect

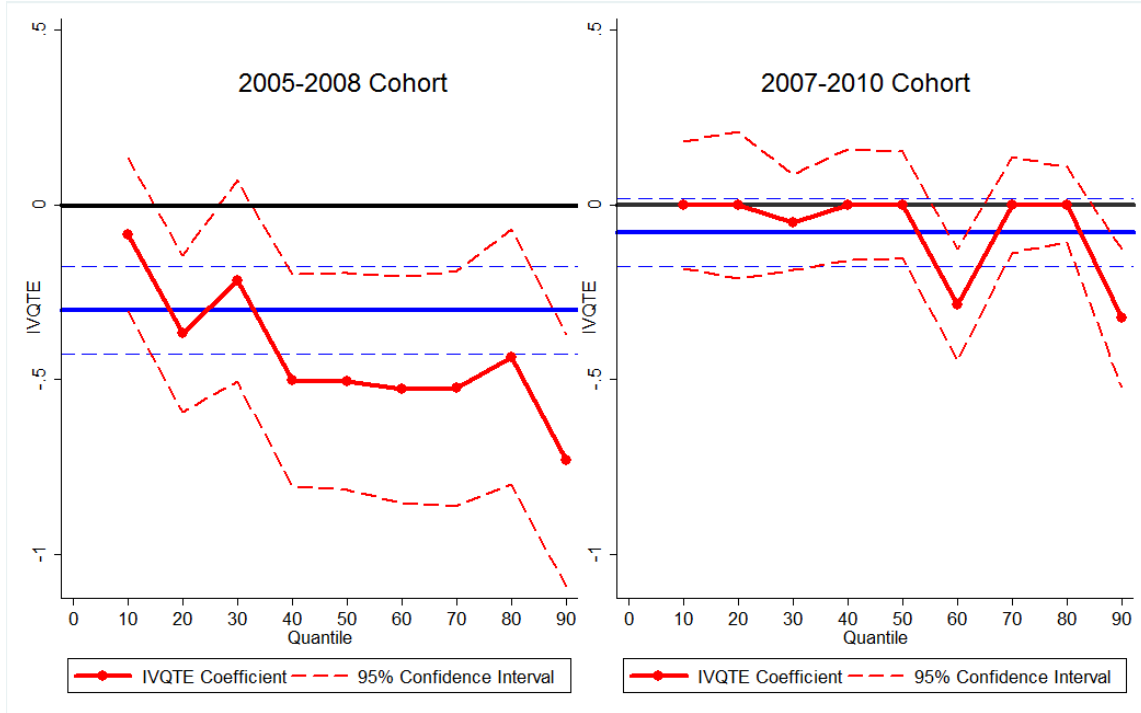
Dependent Variable: Normalized 9th grade score									
Quantile	10	20	30	40	50	60	70	80	90
2005-2008 cohort	-0.0845 (0.112)	-0.368*** (0.114)	-0.218 (0.147)	-0.501*** (0.154)	-0.505*** (0.158)	-0.528*** (0.165)	-0.525*** (0.171)	-0.435** (0.186)	-0.731*** (0.184)
2007-2010 cohort	0 (0.0926)	0 (0.107)	-0.0496 (0.0698)	0 (0.0811)	0 (0.0778)	-0.286*** (0.0811)	0 (0.0698)	0 (0.0550)	-0.324*** (0.0999)

Each cell reports the coefficient of policy school dummy in a separate instrumental QTE regression, which uses losing lottery to instrument for attending a policy school. The coefficient indicates the difference in distribution of normalized 9th grade score of 6th graders who were randomly assigned to a chosen school and a policy school.

Before the policy, seven out of nine coefficients for 2005-2008 cohort were significantly negative, which means that the value-added to students by policy schools is lower for most of distribution. After the policy, all coefficients are indifferent from zero or less negative than before. This shows that the top ten-percent quota policy narrowed value-added gap between the low-performing schools and the over-subscribed schools.



Figure 1.5: Changes in Value-added Gaps



Notes: These two figures plot estimates of 2SLS and instrumental quantile regressions from 10 percentile to 90 percentile with a 10 percentile increment for two cohorts, one before the ten-percent quota policy and one after the policy. 2SLS estimate and its 95% confidence interval are in blue; those for IVQTE analysis are in red. These estimates indicate the value-added gap between a policy school and an over-subscribed school.

The fact that the policy helped narrow the value-added gap for some low-performing and middle-range students provides evidence of peer effects. High-performing students also benefited from the policy in terms of value-added, which could be a mixture of tournament and peer effects. In fact, if we move the 2005-2008 cohort value-added estimates upward and compare that with 2007-2010 cohort's estimates, the top thirty percentiles were moved upward a little more than the other deciles, which may be some suggestive evidence for tournament effect, or may come from a nonlinear peer effect. The instrumental QTE

analysis does not provide strong evidence for the tournament effect, but supports that peer effects are at work.

## 1.6 CONCLUSION

This paper evaluates a current education policy in Changsha that aims at improving low performing middle schools. By guaranteeing seats in elite high schools to the top ten-percent of students attending the policy schools, it seeks to improve the desirability of these schools, attract better incoming students, bring positive peer effects and encourage students to compete for the top ten-percent. I document that the policy helped narrow the gap between low-performing policy schools and the other schools. Sixth graders with better math scores and socioeconomic status are more likely to voluntarily enroll in the policy schools, which improved the incoming students' quality. Instrumental variable estimation shows that the value-added gap between policy schools and over-subscribed schools was closed, and instrumental quantile treatment effect estimates further confirms the effect is prevalent across different deciles. It suggests that the top ten-percent quota policy was successful in equalizing the performance and school quality between the low-performing schools and over-subscribed schools.

There are two limitations of the study. First, this study cannot perfectly disentangle different channels influencing changes in value-added, including peer effects, tournament effects and possible teacher behavioral changes. Second, I do not observe the general equilibrium effect, in other words, overall impact on the whole city. Instead of improving low-performing school while keeping over-subscribed school quality the same, it might be true that the previous performance gap and value-added gap were all driven by peer composition; thus changes in sorting patterns closed both gaps. In future work, I plan to collect college entrance exam performances for different counties in Hunan province across

years to analyze the general equilibrium effects.

Results here on the top ten-percent quota policy complement findings in top ten-percent rule in Texas by [Cullen et al. \(2013\)](#). Students do respond to the incentive of relative grading. This study advances previous findings by estimating the value-added gap before and after the policy introduction, given the unique lottery school assignment system. To what extent this result would apply to the environment with larger variation in racial composition and instruction is unknown. The less racial difference inhibits peer effects, the more this study could speak to cases such as Texas.

The study provides implications on how a government mandate on school admission process influence sorting behavior, student outcomes and school outcomes. Early in 2014, the Chinese Ministry of Education demanded local governments to abandon merit-based admission and instead, enforce proximity-based admission that may be combined with choice-based lottery school assignment. The aim is to curb sorting and equalize access to school resources for students across the socioeconomic spectrum. This study, as well as previous studies, has shown that students with low socioeconomic status are less likely to choose a good school, which suggests that school-choice program is not a panacea and may not suffice to close the income achievement gap. Policies that change sorting patterns and attempt to improve low-performing schools' quality are worth further exploration.

## 2.0 PEER MENTORING AND GROUP INCENTIVES: EVIDENCE FROM CHINESE RURAL MIDDLE SCHOOLS

### 2.1 INTRODUCTION

Peer-assisted learning strategies have long been used and proven to be a cost-effective intervention ([Levin et al., 1984](#)), yielding gains in both academic and transferable social and communication skills to all participants ([Cohen et al., 1982](#); [Rohrbeck et al., 2003](#)).<sup>1</sup> Previous studies have used field experiments to evaluate the effects of the peer tutoring/group study interventions ([Angrist et al., 2009](#); [Li et al., 2014](#); [Blimpo, 2014](#)). Researchers have found evidence that incentives can sometimes be effective for encouraging and improving learning, but that monetary incentive tied to test scores may not work because students do not know how to improve learning by themselves ([Kremer et al., 2009](#); [Bettinger, 2012](#); [Fryer, 2011](#)).

Low education quality and high drop-out rates are problems prevalent in the low-income rural regions of developing countries. In rural Northwestern China, more than 14% of middle school students drop out during their Nine Year Compulsory Education ([Yi et al., 2012](#)). Given the mixture of educational and social/emotional issues faced by students, many of whom live away from their families while attending school, peer mentoring may address the diverse needs of mentees, while also benefiting mentors, who could derive

---

<sup>1</sup>Classwide peer tutoring is a well-known type of peer-assisted learning, in which students in a class are paired to work together. Research indicates that it can significantly improve students' performance in reading, spelling, and math ([Fuchs et al., 1997](#); [Fantuzzo et al., 1992](#); [Greenwood et al., 1989](#)).

pleasure and pride from helping their peers and learn from teaching.

We developed a peer mentoring program, which pairs high-achieving students as mentors to lower performing classmates of the same gender<sup>2</sup>. Incentives include snacks during peer mentoring time (input) and prizes to award an increase in pair ranking change (output). Previous research has documented benefits from introducing team-based incentive for education, especially for disadvantaged groups (Li et al., 2014; Blimpo, 2014).

We pilot-tested the program with two rural middle schools in Yulin, Shaanxi Province during the fall semester of 2013. We targeted the peer mentoring program at 7th grade students at No. 5 Middle School and 8th grade students at Liangzhen Middle School; 8th grade students at No.5 Middle School and 7th grade students at Liangzhen Middle School comprised the control group. In total, there are six classes in each group, with 242 students in the treatment group and 216 students in the control group. We conducted demographic and mental health surveys and administered standardized math exams both before and after the intervention.

The peer mentoring program had both intended and unintended effects. On the positive side, mentors experienced a 0.56 standard deviation improvement in math score. In addition, none of the mentors in the treatment group dropped out of school during the semester, and among the high performing students in the control group (who would have been mentors had they been in the treatment group), only three out of 99 students dropped out. The program also helped decrease the social stress and absence rates of mentors by once per year, which is 28.6% lower than the control group.

However, contrary to the program's intention, mentees exhibited no gains in math scores and a 0.2 standard deviation decline in mental health scores after participating in the program. Breaking the mental health scores into subcategories, we found that the worsened mental health scores were mainly due to higher learning stress.

Even though objective evaluations showed that mentors benefited the most, mentees

---

<sup>2</sup>Rohrbeck et al. (2003) found that same-gender peer-assisted learning studies have a greater effect size on average than mixed-gender ones.

reported a higher level of program helpfulness than mentors: 55 percent of mentees thought that the peer mentoring program was very helpful whereas only 28 percent of mentors did so. In addition, mentees were more likely to report that they would choose the same group partner again (84%) than mentors (54%).

Our study is related to prior research examining the effect of peer mentoring/ tutoring (Topping, 2005; Angrist et al., 2009), incentives to learn (Kremer et al., 2009), and education interventions in developing countries (Kremer and Holla, 2009). In most of previous peer mentoring/tutoring studies, the focus is the impact on mentees. When mentors were studied, it was mainly to examine their fulfillment from responsibilities. Those studies that did examine academic impacts on mentors primarily used self-reporting, which our study shows can be unrelated to objective measures. Our study provides what we believe is the strongest evidence to date of positive effects on mentors' achievement.

We also found suggestive evidence that frequent and immediate reward for inputs (ie. time) attracted more attention and provided greater motivation than future and uncertain rewards for team performance. Unfortunately, we cannot discern whether this was because frequent and immediate reward is more salient and effective or because incentives for input work better than incentives for outcomes (Fryer, 2011).

Lastly, we tested, in the setting of a developing economy, an educational intervention which is easy to implement and scale because of its simple design and low cost. Compared to previous studies on education interventions in developing countries, such as computer-assisted learning, teacher financial incentives, remedial and contract teachers, merit scholarships, or conditional cash transfers, the impact on mentors (0.56 standard deviation) in our pilot study is among the highest (Kremer et al., 2009). The other nice feature of the program is that the per capita cost is very low: less than \$10 USD per treated student per semester. The cost could be even lower if we decreased the frequency of snack prizes from daily to weekly. Given the feedback from students, a weekly snack prize would probably still suffice as motivation for peer mentoring input.

The rest of the paper is organized as follows. Section 2 describes the design of the peer mentoring program. Section 3 discusses pilot trial results and feedback. Section 4 discusses possible factors that drive our results and proposes program modifications that could potentially prevent the unintended effects on mentees.

## 2.2 PROGRAM DESIGN

The peer mentoring program matched top students as mentors to lower-performing students, and provided incentives for the two to study together, improve their academic performance and complete their education. Mentors can work with mentees to study class materials, discuss homework questions, review the questions they got wrong on previous homework assignments and exams, highlight the key points for the exams, provide encouragement and support, and become a role model. Mentees would ideally better understand the class materials and feel supported and encouraged. Witnessing their own improvement, they would ideally have more confidence and interest in their studies, leading to more effort, better outcomes and less likelihood of dropping out.

Mentors could also benefit from the program, consistent with the saying that “to teach is to learn twice.” Giving and receiving elaborated help and providing answers is associated with mathematics achievement ([Webb and Farivar, 1994](#)). [Greenwood et al. \(1989\)](#) shows that mentors experience greater achievement than mentees. Mentors could also, ideally, experience a sense of accomplishment from their mentees’ progress.

### 2.2.1 Pairing method

To create mentor/mentee pairs, students in each class were segregated by gender and ranked according to their overall academic performance from the previous semester (based on their performance on reading, math, English, politics, history, geology and biology),

and divided into two equal size groups of those in the top and bottom 50th percentiles. We then matched same-sex pairs of students within their percentile group. For example, the top female student was matched with the female student who was just below the 50th percentile within female students. If there were 40 students in a class (20 boys and 20 girls), the girl ranked first would mentor the 11th ranked girl, and the 10th ranked girl would mentor the 20th ranked girl. The ranking gap between the mentors and mentees is, by design, the same for every pair. We arranged the pairing in this way to maximize the difference and consistency of the difference in academic performance between the mentor and mentee.

### **2.2.2 Group Incentive Design**

The peer mentoring program incentivized students to spend time together (input) and improve the pair's class rank (output). To encourage the pairs to study together, we provided tasty, healthy snacks (like rice cake, dairy candies or fiber bars) for class teachers to reward both mentors and mentees if they studied together. They could study between classes, at morning and night self-study sessions, or during lunch and dinner breaks. Students were asked to take notes so that the class teacher could verify the interaction between the pair and reward accordingly with snacks. The pairs who did not take notes or randomly copied something onto the notebook would not get a snack from the class teacher. Pairs who studied together four or more times each week received a bonus snack at the end of the week.

In addition, rewards were available to pairs based on the improvement in the average ranking of mentor and the mentee in the school's monthly exams that cover all subjects. We used a comprehensive exam so that students would not be incentivized to put more effort into only one subject (if performance on one were incentivized) at the expense of others. We chose to use the schools' own exams to encourage the students to closely follow and grasp what they were taught in school. This design also makes the incentivized peer



mentoring program more scalable because it does not require testing beyond what a school already conducts.

For midterm and final exams, students were rewarded according to the following procedure. First, we rank the pairs by improvement from their previous ranking, which is the average rank of both mentor and mentee. Then we divide them into three tiers - a top third, middle third and bottom third, with the top third consisting of most improved pairs. Each month, all students receive a surprise prize, but the top third of improvers get a slightly superior prize to the middle third, and the middle third receive a slightly better prize than the bottom third. Prizes, such as notebooks, pens or pencils, were diverse enough to differ in desirability. As an example, if a mentor improves his ranking by 1 and a mentee improves his ranking by 3, the pair's average rank would increase by 2. If a first tier pair's ranking improves by 2 to 5 ranks, this pair would get the best type of prize.

The outcome-based incentive scheme was designed to incorporate several features. First, the fact that everyone gets a prize, regardless of his or her performance, should maintain student participation and help foster a positive attitude toward the program. Second, since the reward is based on the improvement in ranking, all students, whatever their rank, have the opportunity to earn rewards. Third, it is unlikely that the same pair would improve or decline in average ranking each time, so different pairs are likely to get top prizes each month. However, if either member of the pair drops out, no reward will be given. This provides students with the incentive to provide social support to prevent their partner from dropping out of school.

### **2.2.3 Implementation**

We implemented the peer mentoring program in the fall semester of 2013 in the 7th grade at No. 5 Middle School and 8th grade at Liangzhen Middle School, and used the 8th grade at No.5 Middle School and 7th grade at Liangzhen Middle School as control groups. There are six classes in each group, with a total of 242 students in the treatment group and 216

students in the control group.

Our research group conducted baseline surveys and standardized math tests in late September 2013. Research assistants located in the area explained the peer mentoring program to teachers and students and announced the group matching according to the previous semester's overall ranking. After the initial visit, the researchers visited the schools twice during the fall semester to give the teachers snack rewards and prizes for improvement. We conducted an evaluation survey and math test in late February 2014. The expenses for snacks, prizes for improvement, and compensation for extra work put in by the teacher to check the notes and give snack prizes equaled approximately \$2,100, \$150, and \$100 USD, respectively, totaling to approximately \$2,350 USD.

#### **2.2.4 Balance Check: Pre-Treatment Characteristics**

The baseline surveys include three main parts: 1. A student survey with questions on family backgrounds, attitudes towards school, self-efficacy and emotion management, predictions about who was most likely to drop out in class, etc. A mental health survey was included in the baseline survey. We computed the mental health score based on answers to ninety questions on learning stress, fear, impulsiveness, etc..

2. A standardized math test with twenty-five multiple-choice questions.

3. A teacher survey asking about their qualification, effort, values and predictions about which students were most likely to drop out.

In this subsection, we look at baseline survey and math test scores to gain a general picture of our sample and conduct a balance check to see how mentors and mentees differ from their counterparts in the control group. [Table 2.1](#) presents selected baseline characteristics of mentors and mentees in treatment and control groups. Most of the pre-treatment characteristics are balanced between two groups, including mental health score, parental education, family size, poverty program participation, and classroom behavior. However, 7th and 8th grade mentees in the treated classes have baseline math scores that are 0.21

and 0.67 standard deviation higher than those in the control group. The class size is on average 4 students larger (40.5 versus 36.5) and class participation per day is about 1.5 times less for treatment group students, which may bias against a positive treatment effect.

Other than the aforementioned, we did extra balance checks and present the results in [Table A6](#). Mentees reported a 0.7 lower rating of “like going to school” on a scale from 1 to 10 relative to their counterparts in the control group, which may also create bias against a positive treatment effect on academic and mental health improvement. Mentees in the treatment group reported arguing with classmates less frequently and mentors in the treatment group reported fighting with classmates less frequently, compared to the control group. Note that there were 42 different comparisons (mentor and mentee \* 21 items), so we would expect about 2 significant differences on the basis of chance; however, there were 10, albeit not going in any systematic direction in terms of favoring the treatment or control group.

There was no statistical significance between treatment and control groups for class teachers for the following characteristics: years of teaching experience, public qualification, years of obtaining the qualification, years being the class teacher, or any other measured attribute (see Appendix for details). Since we do not have perfectly balanced baseline characteristics between treatment and control groups, especially for mentees, we include baseline controls in all regressions with an outcome as dependent variable in our quantitative evaluation.

### **2.3 EVALUATING THE PEER MENTORING PROGRAM**

We analyze the treatment effects on mentors’ and mentees’ math scores, mental health survey scores, staying in school or not, and other outcomes such as school attendance, study habits and nonacademic behavior. We also summarize the subjective evaluations

Table 2.1: Baseline Student Characteristics and Balance Check

	Mentee			Mentor		
	Control	Treatment	Diff	Control	Treatment	Diff
Female	0.47	0.52	0.0502	0.475	0.5	0.0253
	-0.501	-0.502	-0.0648	-0.502	-0.502	-0.0684
Class size	36.47	40.468	<b>3.998***</b>	36.556	40.559	<b>4.004***</b>
	-4.448	-2.676	-0.476	-4.529	-2.652	-0.516
Grade 7 Std Math	-0.689	-0.476	<b>0.213*</b>	0.633	0.76	0.127
	-0.452	-0.748	-0.124	-0.765	-1.119	-0.198
Grade 8 Std Math	-0.864	-0.224	<b>0.640***</b>	0.488	0.504	0.0156
	-0.414	-0.804	-0.109	-0.773	-1.104	-0.178
Mental Health	61.63	59.25	-2.375	62.25	61.84	-0.409
	-14.176	-14.129	-1.888	-14.667	-12.808	-1.946
Poverty Program	0.353	0.315	-0.0389	0.404	0.333	-0.0707
	-0.48	-0.466	-0.105	-0.493	-0.473	-0.0998
Father Education	2.094	2	-0.094	2.202	2.246	0.0437
	-1.008	-1.004	-0.154	-1.078	-1.054	-0.188
Mother Education	1.701	1.782	0.0814	1.808	1.822	0.14
	-0.94	-0.984	-0.182	-0.986	-0.957	-0.141
Absence	1.077	1.149	0.0723	1.747	1.822	0.102
	-0.683	-0.723	-0.13	-0.747	-0.833	-0.162
Obs	117	124		99	118	

Note: First three columns takes the lower half performing students based on baseline math score in the control classes as the control group for mentees. Similarly for mentors. “Control” and “Treatment” column show the means and standard deviations in parentheses. “Diff” column shows the difference and robust standard errors in parentheses, with significance level indicated by \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . Baseline math scores are standardized within grade. Mental health score is calculated based on 100 questions, with higher score meaning better mental health condition. Parental education is categorical: less than elementary (1), elementary (2), middle school (3), high school (4), vocational college (5), college graduates or above (6). Absence is self-reported frequency per semester.

and suggestions for the program, both of which provided by students and teachers in the treated classes.

### 2.3.1 Main Results

We focus on three categories of outcomes, math score, mental health, and drop out rate, in our main results. We hoped to see higher standardized math exam performance, better mental health survey scores, and lower probability of dropping out (albeit possibly only in the long run).

The baseline regression we use is:

$$Y_{isgc} = \beta_0 + \beta_1 T_{sg} + \beta_2 X_{isgc} + \epsilon_{sgc} + \mu_{isgc} \quad (2.1)$$

where  $Y_{isgc}$  are a set of outcome variables, including post-intervention standardized math score, mental health score, and a dummy for dropping out. When the outcome variable is dichotomous, we use probit model for the estimation.  $T_{sg}$  is a dummy variable for program treatment status, which equals to one if grade  $g$  of school  $s$  is treated.  $X_{isgc}$  represents a set of pre-program characteristics, including standardized math score, gender, grade level, parental education level, etc.. Since we only have twelve classes in this experiment, we used wild cluster bootstrapping method from [Cameron et al. \(2008\)](#). This specification is used to produce [Table 2.2](#), [Table 2.3](#) and [Table 2.4](#), with different sub-samples and outcome variables.

To evaluate treatment effects separately for mentors and mentees, we use the upper and lower half ranking students in the control group as the counter-factual groups for mentors and mentees. More specifically, we assign a placebo mentor or mentee identity for each student in the control group: those who scored above median among the students of the same gender in class are assigned as placebo mentors; those who scored below are placebo mentees.

[Table 2.2](#) shows treatment effects for post standardized math score, post mental health

score, and drop-out probability for mentees (Panel A ) and mentors (Panel B) separately. Treatment effect on alternative outcome measures, changes in pre and post math scores and mental health scores, are analyzed and reported in column (2) and (4).

Panel A reports a negative impact on the change in standardized math scores for mentees. It also shows that mentees in the treatment group report have a significantly worse mental health score than the control group, with a size of 0.2 standard deviation. Columns (7) and (8) show that the peer mentoring intervention reduces dropout probability by around 0.81%, although the coefficients are not statistically significant.

Panel B, in contrast, shows that mentors experienced gains in both math score and drop out rate, and no change in mental health scores. They had a 0.56 standard deviation of improvement in math score with no significant change in mental health. In addition, mentors reported 0.5 times fewer absences per semester, which is approximately 28.6% fewer compared to the control group. No mentor dropped out of school during that semester in the treatment group, while three students ranked in the top half of the control group did.

### **2.3.2 Treatment Effects on Subcategories of Mental Health Survey, Study Attitude, Effort and Behavior**

Using the comprehensive survey, we estimate treatment effects on a larger set of outcome variables using [Equation 2.1](#) and present the results in [Table 2.3](#) and [Table 2.4](#) for mentees and mentors, respectively. Each coefficient reported comes from a separate regression that includes grade and gender controls. Errors are clustered at the class level. Two specifications are used: one uses post scores as the dependent variable and includes baseline scores as control; the other uses change in score (post-baseline) as the dependent variable and does not include baseline in control. Most results reported for these two specifications are similar. Discrepancies happen when the baseline characteristics are not balanced between control and treatment groups. In those cases, we look at the results from the second

Table 2.2: Treatment Effect on Math, Overall Mental Health and Dropout Probability

Model	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) Probit	(8) Probit
Dependent Variable	Post math	std Diff math	Post mental health	Diff men- tal health	Post absence	Diff absence	Dropout	Dropout any
<b>Panel A: Mentee and Placebo Mentee Subsample</b>								
<b>treat</b>	-0.145	-0.405**	-3.147*	-2.953	-0.111	-0.272	-0.132	-0.128
	-0.146	-0.156	-1.702	-1.84	-0.146	-0.22	-0.282	-0.331
Obs	224	224	200	200	224	224	240	240
R-squared	0.079	0.054	0.498	0.031	0.041	0.029		
<b>Panel B: Mentor and Placebo Mentor Subsample</b>								
<b>treat</b>	0.570**	0.564**	-0.128	0.164	-0.478***	-0.576**	na	na
	-0.27	-0.278	-0.723	-1.594	-0.145	-0.239		
Obs	209	209	190	190	209	209		
R-squared	0.168	0.064	0.59	0.032	0.168	0.069		

Significance level \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Standardized Math Score has a mean of 0 and standard deviation of 1 for each grade in each exam. Mental health score is calculated according to a 100 question survey, with higher score being better mental health condition. Absence is student self-reported frequency of absence per semester. Dropout in column (7) equals to 1 if we know a student dropped out of school. The dependent variable dropout any in column (8) equals to 1 if a student dropped out or if he/she is missing from our sample without knowing the reason.

All regressions have gender, grade and baseline controls. Columns (1)-(6) are OLS regression results with robust standard errors clustered at class level in parentheses, corrected for small number of clusters using wild cluster bootstrapping with 1000 replications. (7)-(8) are probit regression results with robust standard errors clustered at class level in parentheses. Column 7 and 8 are empty in Panel B because no mentor in treatment group dropped out.

specification.

Panel A in [Table 2.3](#) and [Table 2.4](#) separately present treatment effects on subcategories of the mental health score on mentees and mentors. By decomposing the total mental health score into subcategories, we see in Panel A in [Table 2.3](#) that the worsened mental health scores in mentees is mainly driven by higher learning stress. More daily study time and the responsibility to improve performance and ranking in return for their mentors' help may have led to the higher learning stress of mentees.

Panel A in [Table 2.4](#) shows that mentors experienced less social stress and were less anti-social after the peer mentoring program. The interactions within pairs may have helped mentors to relieve some social pressure and improve their social skills. However, mentors showed higher level of guilt, which may have resulted from an inability to help resolve mentees' problems. The positive and negative effects of these three subcategories cancel out, which gives us an insignificant change in overall mental health scores for mentors.

Panel B and C present the treatment effects on students' study attitude, effort, and behavior. Contrary to our intention to encourage more questions and less arguments, mentees reported a lower likelihood of asking questions and an increased likelihood of getting into arguments compared with the control group. The peer mentoring program increased mentors' affection for going to school ("On a scale of 1 to 10, how much do you like going to school?"), decreased the frequency of being bullied, and (recall from the main results) lowered absence rate for class. The snack prize may have added more motivation for going to school every day and made mentors like going to school better.



Table 2.3: Treatment Effect on other variables of possible interests  
Mentee and Placebo Mentee Subsample

Depend. Variable	Control Base- line	Panel A: Treatment Effect on Itemized MHT score							
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Learning stress	Social stress	Anti-social	Self-guilty	Over-sensitive	Physical-signs	Fear	Impulsive-ness
Post	Y	0.885*	0.0481	0.387	0.619	0.0788	0.187	0.104	0.195
		-0.479	-0.258	-0.383	-0.441	-0.32	-0.371	-0.37	-0.263
Diff	N	0.873	-0.148	0.469	0.687	-0.0931	0.215	0.0835	0.137
		-0.574	-0.268	-0.378	-0.431	-0.235	-0.518	-0.224	-0.367
Depend. Variable	Control Base- line	Panel B: Treatment Effect on Study Attitude							
		(1)	(2)	(3)	(4)	(5)			
		like school (1-10)	class participation	ask question	read	extra exercise			
Post	Y	-0.203	0.175	-0.118	-0.106	0.0676			
		-0.601	-0.469	-0.105	-0.113	-0.109			
Diff	N	0.0246	0.642	-0.196*	-0.0669	0.0952			
		-0.272	-0.941	-0.1	-0.0888	-0.0936			
Depend. Variable	Control Base- line	Panel C: Treatment Effect on Behavior							
		(1)	(2)	(3)	(4)	(5)			
		bullied	argument	fight	late hw	late class			
Post	Y	0.114	0.521	-0.158	0.0452	0.0138			
		-0.0849	-0.435	-0.163	-0.32	-0.269			
Diff	N	0.0636	1.748*	0.0967	-0.28	0.132			
		-0.106	-0.894	-0.363	-0.379	-0.274			

Note: Robust standard errors clustered at the class level in parentheses, corrected for small number of clusters using wild cluster bootstrapping with 1000 replications. This table only looks at the subsample of mentees in the treatment group and the lower half performing students (based on baseline math score) in the control group, with 225 observations.

All regressions include grade and gender as controls. Reported are coefficients for treatment dummies. For each panel, first row uses post outcomes as dependent variables and include baseline control and second row uses change as outcome (post-base) without baseline control. In Panel A, a positive coefficient means a worse problem in that category. For example, learning stress is higher in the treatment group than the control group. In Panel B and C, “class participation” asks for daily frequency; “ask question”, “read”, “extra exercise” and “bullied” are categorical variables from never (0), to sometimes (1), to often (2); other than “bullied”, all other variables in Panel C are frequency per semester.

Table 2.4: Treatment Effect on other variables of possible interests  
Mentor and Placebo Mentor Subsample

Panel A: Treatment Effect on Itemized MHT score									
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Depend. Variable	Control Base-line	Learning stress	Social stress	Anti-social	Self-guilty	Over-sensitive	Physical-signs	Fear	Impulsive-ness
Post	Y	0.322	-0.550**	-0.579**	0.851*	-0.0729	-0.0969	0.129	-0.203
		-0.283	-0.255	-0.229	-0.457	-0.186	-0.206	-0.623	-0.216
Diff	N	0.508	-0.664**	-0.718**	0.743*	-0.23	-0.0669	0.174	-0.0531
		-0.336	-0.304	-0.306	-0.444	-0.245	-0.194	-0.467	-0.324
Panel B: Treatment Effect on Study Attitude									
		(1)	(2)	(3)	(4)	(5)			
Depend. Variable	Control Base-line	like school (1-10)	class participation	ask question	read	extra exercise			
Post	Y	0.488*	-1.022**	0.119	0.153	0.0741			
		-0.295	-0.509	-0.0953	-0.107	-0.0676			
Diff	N	0.292	-0.139	0.102	0.228	0.0168			
		-0.368	-1.537	-0.0991	-0.161	-0.0996			
Panel C: Treatment Effect on Behavior									
		(1)	(2)	(3)	(4)	(5)			
Depend. Variable	Control Base-line	bullied	argument	fight	late hw	late class			
Post	Y	-0.0812**	-1.979	-0.597	-0.0235	-0.0786			
		-0.0372	-2.148	-0.782	-0.271	-0.288			
Diff	N	-0.0278	-1.621	-0.355	-0.208	0.074			
		-0.0942	-2.655	-0.488	-0.369	-0.298			

Note: Robust standard errors clustered at the class level in parentheses, corrected for small number of clusters using wild cluster bootstrapping with 1000 replications. This table only looks at the subsample of mentors in the treatment group and the top half performing students (based on baseline math score) in the control group, with 209 observations.

All regressions include grade and gender as controls. Reported are coefficients for treatment dummies. For each panel, first row uses post outcomes as dependent variables and include baseline control and second row uses change as outcome (post-base) without baseline control. In Panel A, a positive coefficient means a worse problem in that category. For example, learning stress is higher in the treatment group than the control group. In Panel B and C, “class participation” asks for daily frequency; “ask question”, “read”, “extra exercise” and “bullied” are categorical variables from never (0), to sometimes (1), to often (2); other than “bullied”, all other variables in Panel C are frequency per semester.

Figure 2.1: Kernel Density Plots of Mentees' Standardized Math Scores

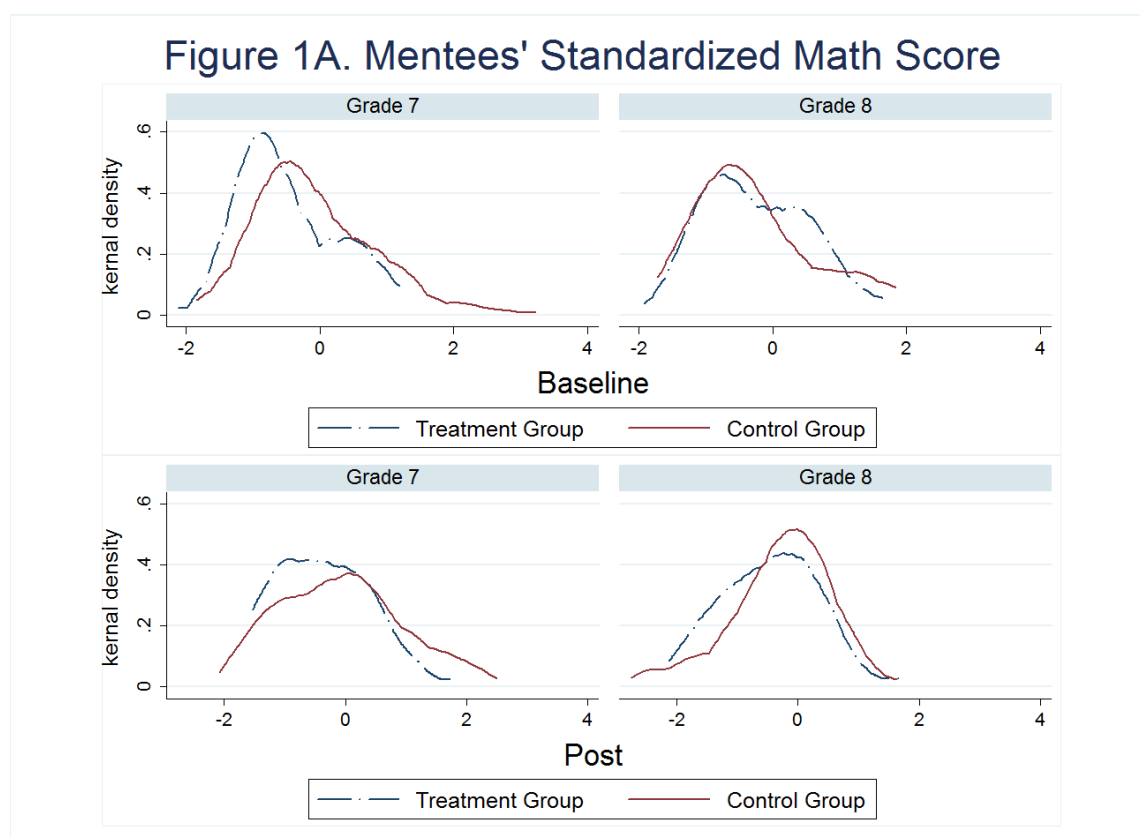


Figure 2.2: Kernel Density Plots of Mentors' Standardized Math Scores

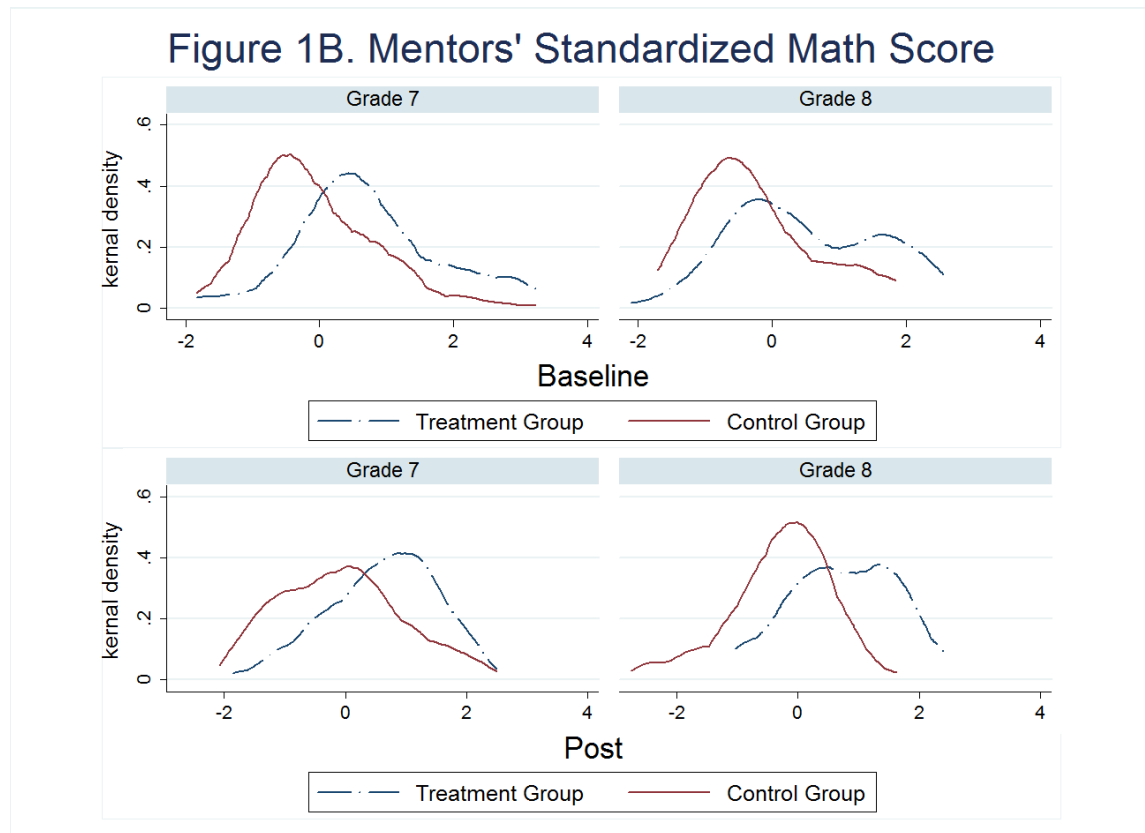
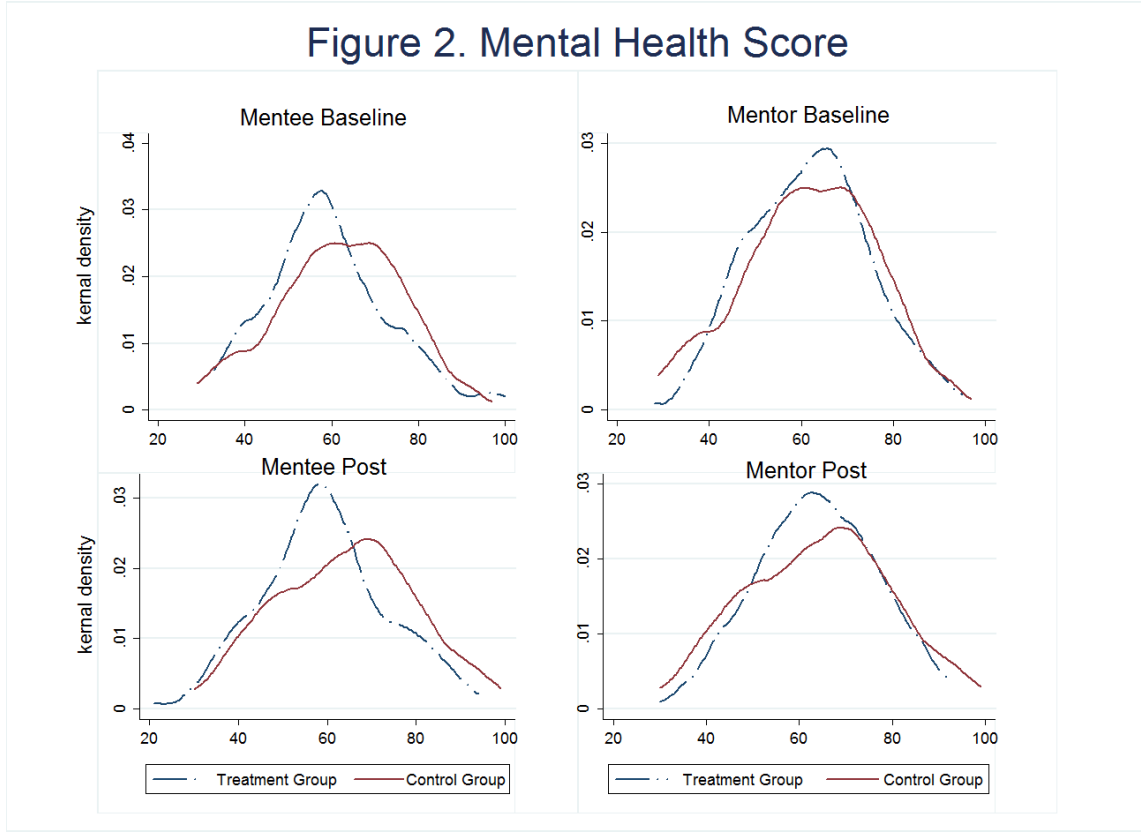


Figure 2.3: Kernel Density Plots of Mental Health Scores



### 2.3.3 Heterogeneous Treatment Effects

Figure 2.1, Figure 2.2, and Figure 2.3 show the kernel density plots of math scores and mental health scores pre and post-treatment. We separately plot the math scores of 7th and 8th grade, because they took different exams and scores are normalized separately. The observations on shifts in distributions are consistent with the results in Table 2. They show that the negative coefficients on mentees' math and mental health score come from

positive changes in the control group while the treated mentees stayed the same.

In [Figure 2.1](#), standardized math score distributions did not change much for mentees in the treatment group, but shifted to the right for lower performing students in the control group. Two plots on the left side of [Figure 2.3](#) tell a similar story: there is a rightward shift in mental health score distribution for mentees in the control group while the treated mentees stayed the same. Meanwhile the distributions of mentors' standardized math scores shifted rightward significantly, as shown in [Figure 2.2](#). [Figure 2.3](#) provides little evidence of a systematic shift in mental health score distributions for mentors.

We further quantify the treatment effect across students with different academic performance and mental health with quantile regressions.

$$Y_{isgc} = \alpha_0^p + \alpha_1^p T_{sg} + \alpha_2^p X_{isgc} + \eta_{isgc}^p \quad (2.2)$$

where  $0 < p < 1$  indicates the proportion of the population having scores below the  $p$ th percentile.

[Figure 2.4](#) and [Table 2.5](#) presents quantile treatment effects on math scores, for both mentees and mentors. [Figure 2.4](#) shows that the peer mentoring program improved standardized math score for almost all mentors with no significant improvement (there were, in fact, negative point estimates at all quantiles except the lowest) for mentees. Mentors generally gain more from the program when they have better academic performance, although the top 10% students may not benefit as much. The gains in post standardized math score increased for mentors from 10 to 75 percentile, but decreased from 0.783 standard deviation gain for mentors with standardized math scores at 75 percentile to 0.492 standard deviation for those at 90 percentile.

Figure 2.4: Quantile Treatment Effects on Standardized Math Scores

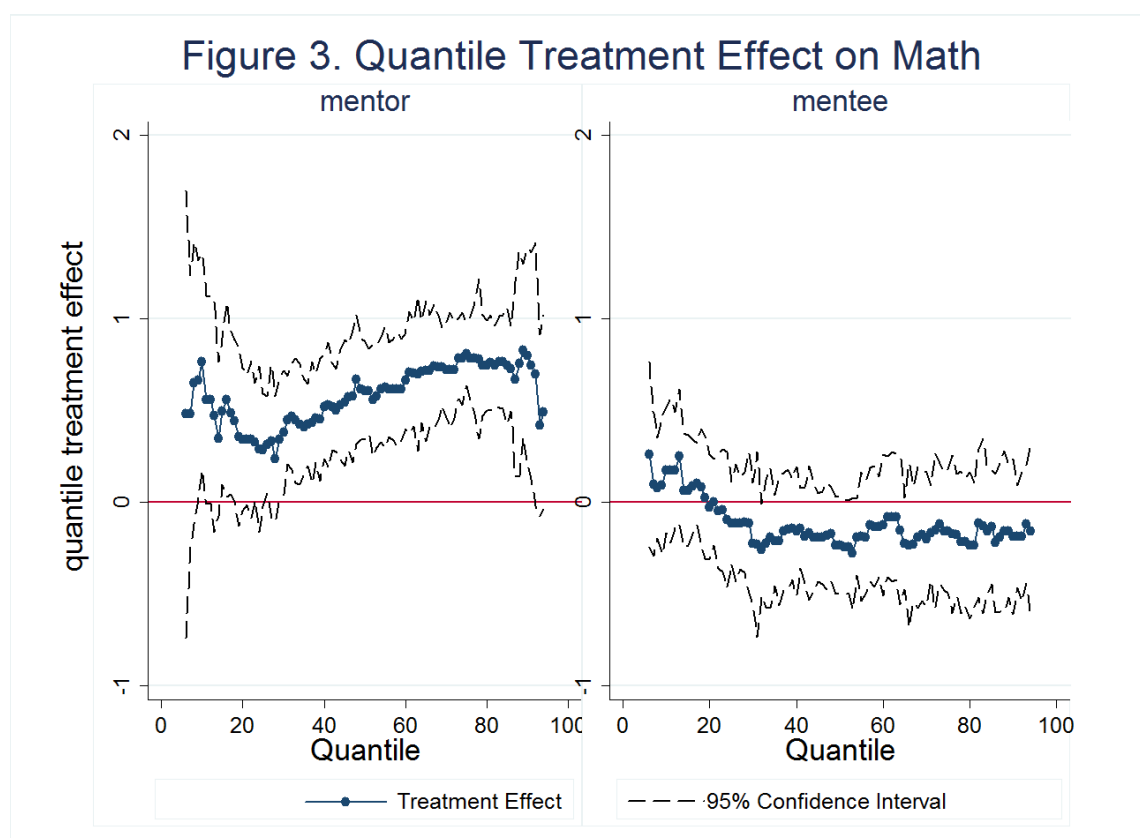


Figure 2.5: Quantile Treatment Effects on Mental Health Scores

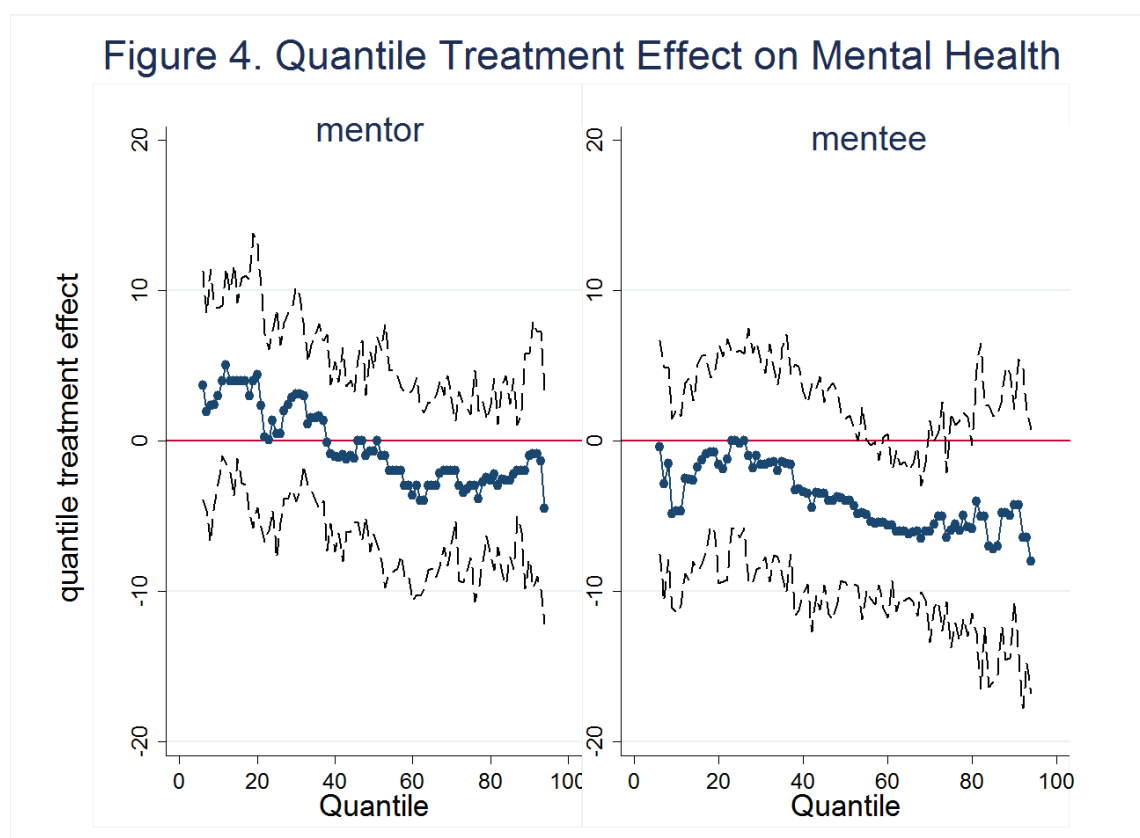




Table 2.5: Heterogeneous Effect on Math Score: Quantile Regressions

<b>Panel A: Mentee and Placebo Mentee Subsample</b>					
Quantile	10%	25%	50%	75%	90%
Treat	0.0656	-0.193	-0.19	-0.157	-0.157
	-0.19	-0.166	-0.155	-0.171	-0.263
Observations	224	224	224	224	224
<b>Panel B: Mentor and Placebo Mentor Subsample</b>					
Quantile	10%	25%	50%	75%	90%
Treat	0.346	0.425**	0.616***	0.783***	0.492
	-0.333	-0.167	-0.151	-0.143	-0.3
Observations	209	209	209	209	209

Notes: All regressions include grade, gender and baseline math scores as control. Significance level indicated by \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . Panel A takes the subsamples of mentees in the treatment group and the lower half performing students (based on baseline math score) in the control group. Similarly for Panel B. Standardized Math Score has a mean of 0 and standard deviation of 1.

The quantile treatment effects on mental health are shown in Figure 4, where we see no significant effect on mentors but negative effect on mentees, mainly for the top 40% in the distribution. It suggests that worsened mental health scores for mentees mainly come from those with better mental health conditions.

Similarly in [Table 2.6](#), mental health scores remain unchanged for mentors but became significantly worse for mentees at the 25th, 50th, and 75th percentile of the mental health score distribution. The magnitude of worsened mental health score is around 0.28 standard

deviations.

Table 2.6: Heterogeneous Effect on Mental Health Score: Quantile Regressions

<b>Panel A: Mentee and Placebo Mentee Subsample</b>					
Quantile	10%	25%	50%	75%	90%
Treat	-1.868	-4.200*	-4.488***	-4.333*	1.5
	-3.529	-2.299	-1.633	-2.333	-2.201
Observations	200	200	200	200	200
<b>Panel B: Mentor and Placebo Mentor Subsample</b>					
Quantile	10%	25%	50%	75%	90%
Treat	-0.517	-1.625	-0.978	2.529	-0.226
	-2.652	-2.033	-1.535	-2.178	-2.498
Observations	190	190	190	190	190

Notes: All regressions include grade, gender and baseline mental health scores as control. Significance level indicated by \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . Panel A takes the subsamples of mentees in the treatment group and the lower half performing students (based on baseline math score) in the control group. Similarly for Panel B. Mental health score is better if higher.

### 2.3.4 Subgroup Treatment Effects

Understanding how the program impacts different subgroups is important for assessing whether it is likely to be more successful in some environments or with some populations. In [Table A8](#), we add to the baseline regression equation (1) interaction terms of the treatment dummy variable with subgroup characteristics.

$$Y_{isgc} = \beta_0 + \beta_1 T_{sg} + \beta_2 X_{isgc} + \beta_3 T_{sg} * X_{isgc} + \epsilon_{sgc} + \mu_{isgc} \quad (2.3)$$

We look at two types of subgroups (gender and grade) and two outcome variables (math score and mental health score). We do not find any significant gender difference in treatment effect on math or mental health.

8th graders contributed the most improvement to math score for mentors. They experience a 0.654 standard deviation greater gain in standardized math score than 7th grade mentors. Moreover, 8th grade mentees experience a smaller negative effect in mental health scores than 7th grade mentees. It is possible that 8th graders were more familiar with each other and could form stronger peer mentoring groups. However, since we treated two 7th grade classes in one school and four 8th grade classes in the other, we do not know if the heterogeneous treatment effects come from the difference in grade levels or the difference in schools.

### **2.3.5 Qualitative Measures**

To understand what might be driving the observed effects of the program, we separately looked at students in the treatment group and tested if the following factors were correlated with more positive treatment effect on math score: different levels of engagement and satisfaction in the program (choosing the same partner again, feeling hopeful, feeling anxious, deeming the program helpful, etc.), peer baseline academic performance, difference between self and peer baseline academic performance, whether the class teacher teaches math, teacher's years of experience, etc. However, after controlling for baseline score and a few characteristics, none of these factors were statistically significantly correlated with post standardized math scores.

Other than the demographic, behavioral and mental health questions we asked in the baseline survey, we also added peer program evaluation questions in the post survey. Overall, more than 90% of students found the program to be at least a little helpful (see [Table 2.7](#)). Not reported in the table, 69.70% of students reported that they would choose the same partner if they were to do the program again. Interestingly, given the largely

positive results for mentors and negative results for mentees, more mentees than mentors reported that the program is very helpful (55% versus 28%) and that they would choose the current partner again (84% versus 54%).

We asked five questions about feelings towards the peer mentoring program, including pride, anxiety, embarrassment, hope and frustration. Mostly the answers were positive. Few students reported that they felt “very” anxious, embarrassed and frustrated. We checked the students who gave very negative answers, such as feeling very anxious/very embarrassed/no hope at all/very frustrated, and found that most of them were low performing students (in terms of standardized math score).

In the evaluation survey, students also reported on mentoring timing, frequency, length and main subject discussed. Most students (93.5%) reported having participated in peer mentoring more than four times a week, the minimum requirement to get the weekly bonus snack. Most peer mentoring activities happened in between classes or during morning and night self-study sessions, and usually for 20-30 minutes per day (73%). We tested whether the frequency and length can help explain the benefit in the academic or mental health score changes of mentor/mentee and did not find any statistically significant evidence.

## 2.4 CONCLUSIONS

The most striking finding in the study is that, contrary to expectations and intentions, mentors gained more from the program than did mentees, who, in fact, mainly suffered negative effects. Mentors benefited on multiple dimensions (better math scores, less social stress, lower dropout rate, more affection for school, less absence, and extra reading after class) with slightly more guilt, which may stem from their inability to always resolve questions from mentees. Mentees did not gain in standardized math score and reported higher learning stress.

Table 2.7: Student Subjective Survey Evaluation

<b>Part A: Is the peer mentoring helpful?</b>					
		mentor		mentee	
		# student	% student	# student	% student
Helpful					
	Not at all	17	14.91	4	3.48
	A little	65	57.02	48	41.74
	Very much	32	28.07	63	54.78
<b>Part B: Does peer mentoring makes you feel __?</b>					
		mentor		mentee	
		# student	% student	# student	% student
proud					
	Not at all	59	51.75	45	39.13
	A little	45	39.47	60	52.17
	Very much	10	8.77	10	8.7
anxious					
	Not at all	63	55.26	60	52.17
	A little	48	42.11	50	43.48
	Very much	3	2.63	5	4.35
embarrassed					
	Not at all	95	83.33	78	67.83
	A little	17	14.91	36	31.3
	Very much	2	1.75	1	0.87
hopeful					
	Not at all	20	17.54	12	10.43
	A little	63	55.26	68	59.13
	Very much	31	27.19	35	30.43
frustrated					
	Not at all	89	78.07	86	74.78
	A little	22	19.3	26	22.61
	Very much	3	2.63	3	2.61

If distributional concerns were not an issue, the program could be deemed a success, since mentor gains were greater than (insignificant) mentee reductions in academic scores. Yet it is unfortunate that the program did not benefit those who needed it the most - the students performing in the lower half.

Although we intended to solicit group incentives for both study effort and academic improvement, the program as implemented turned out to place far more emphasis on snack prize for peer mentoring time (but not quality) than on prizes for group improvement. Mentors and mentees may have spent enough time together, but they studied separately. A student wrote in the post-survey that all they cared about was the snack (, and no one cared about actually helping each other). Other than the possible theory proposed by Fryer (2011), that students do not know the education production function to respond to outcome incentive effectively, it is also possible that students were present-minded and hence more motivated by the immediate and frequent availability of snacks. Finally, it is possible that the quality differences in rewards for improving pair testing rank were not sufficient enough to motivate mentees to improve their scores. The insufficient prize for mentees' improvement may be the main reason why our results are different from previous studies, which found that using monetary prizes as group incentives by itself helps improve test scores (Blimpo, 2010), especially for mentees (Li et al, 2014).

The program may have also increased stigma for those at the bottom, since the pairing mechanism was transparent to students. A cross-age peer mentoring program, which pairs students at higher grades with students in lower grades, might avoid the stigma effect for mentees while still providing students in the higher grades, with the benefits of mentoring. Yet, it is unknown and worthy of study whether lower performing students in the higher grade will also benefit from mentoring, and whether students will benefit from helping others with material that they have learned in previous years.

### 3.0 EFFICIENT PROCRASTINATION

#### 3.1 INTRODUCTION

Procrastination often negatively affects people's study and work, inflicts costs on businesses and governments, and hurts people's physical and financial health ([Ferrari et al., 1995](#); [Tice and Baumeister, 1997](#)). There are three main causes for procrastination. First, people have planning fallacy and are naive with their hyperbolic discounting; second, people lack confidence, and therefore self-handicappers create conditions that make success impossible; third, people have divided inner selves and sometimes bargaining failure occurs ([Akerlof, 1991](#); [O'Donoghue and Rabin, 2001](#); [Frederick et al., 2002](#); [Steel, 2007](#)). The most prevalent solution is to use external tools such as deadlines to help the parts of ourselves that want to work. Studies have shown that these self-committing devices are useful, for example ([Ariely and Wertenbroch, 2002](#)).

People sometimes find procrastinating exciting because the last-minute panic usually sends them off to an intense period of time. They focus all their attention on the goal, ideas flow out of their mind, and their hearts beat fast. At the end, they are satisfied with the work, not just because it satisfies a decent standard (may not be the best they can do), but more importantly, they spend a minimal amount of time on it to achieve that level of quality. If so, these people would postpone their tasks because their prediction of low efficiency of doing things well in advance and high efficiency of doing things right before deadlines and they prefer to suffer for a shorter period of time. However, the problem of

procrastination comes when they procrastinate too much and the time left is not enough to achieve a minimum quality—then they panic and cannot even efficiently use the limited time left.

The under-motivation at the beginning and the over-motivation near the end yields a hump-shape curve relationship between efficiency and timing. Building upon previous research on optimal arousal theory (Yerkes and Dodson, 1908) and attentional focus theory (Karau and Kelly, 1992), this study develops a conceptual framework of efficient procrastination. The theory predicts that people enjoy a high efficiency, thus perform better and spend less time on the task when they work not too early nor too late – “efficiently procrastinate”.

The study provides empirical evidence to support the hypothesis, using online homework tracking data from an undergraduate economics class. I examine the relationship between how early a student starts/submits/works on the homework affects the homework performance and time cost. The within-subject analysis rules out self-selection problems, which plague many previous studies on procrastination. Results show a non-linear relationship as predicted: the normalized homework score is higher and the time cost is lower when he/she submits the homework neither too early nor too late.

The rest of the paper is organized as follows. Section 2 develops the conceptual framework of efficient procrastination. Section 3 describes the source and construction of the data. Section 4 presents the empirical strategies and the results. Section 5 concludes and discusses possible confounds and solutions.

### 3.2 CONCEPTUAL FRAMEWORK

Previous research shows that people prefer to incur a loss immediately rather than delay it. This may seem to go against procrastination because if people prefer to pay the cost



as soon as possible they shouldn't postpone it. However, many people have the experience that they usually concentrate best and do things most efficiently right before the deadline. With a far-away deadline and little pressure, people tend to slow things down and fail to complete the task within the time needed if they focus. On the other hand, people prefer to spend a shorter time period to do work conditional on the same output, which not only saves time for other things but also gives a feeling of fulfillment. Therefore, it is a strategic choice to reasonably procrastinate. For some people, they are overconfident about their ability of completing the task and begin the task too late. Close deadline and heavy pressure may lead to failure of focus, or a "choking" effect (Baumeister, 1984). In that case, instead of procrastination, people may lower the quality of completion and may even give up the task.

This framework is consistent with two existing models, the Yerkes-Dodson Law (Yerkes and Dodson, 1908) and the attentional focus model by Karau and Kelly (1992). The Yerkes-Dodson Law examines variance in habit formation based on the strength of stimuli: when stimuli were either too intense or lacked intensity, a drop in performance occurred. The attentional focus model suggests that time scarcity focuses group members' attention on most relevant cues for task completion therefore improves group performance.

I propose the following framework to incorporate people's preferences on both work quality and time cost, as well as the non-linear progression of work efficiency, which first goes up then goes down as time approaches the deadline.

$$\max_T U(y(T), c(T)) \quad (3.1)$$

where  $U$  is the utility function increasing in the outcome,  $y$ , and decreasing in the time cost,  $c$ .  $y$  and  $c$  are both a function of  $T$ , which is the time available before the deadline.  $y' > 0, y'' < 0, c' < 0$ , and  $c'' > 0$ . Graphical presentations of these two equations are a hump-shape curve of performance and how early a person starts to work (can be viewed as not procrastinating) and a U-shape curve of time cost and working early. However, taking into account that a person also has less time to be taken away when it is very close to the

deadline, we may observe a downward sloping curve of time cost and working early.

At the beginning of the task, people are usually under-motivated because the available time is abundant. Therefore, people are not very concentrated and easily distracted when working, which leads to longer time cost and lower work quality. As time goes by, when one has just enough time to finish the work by the deadline, productivity is at its highest level. One is most concentrated, which results in high efficiency, low time cost, and high work quality. When one starts to work when it is too close to the deadline, however, productivity goes down. In this situation, one is under pressure and overly motivated. Choking effect would prevent him from focusing on the task and make him do worse than usual. One may panic and cannot stop thinking about the aftermath of failure in task completion.

The conceptual framework predicts that given a time-constraint task, if a person works on it at a timing that is not too early nor too late, he/she would focus the most and therefore enjoy a high performance with a low time cost.

### **3.3 DATA**

#### **3.3.1 CourseWeb Tracking Data**

The panel data set used to test the predictions is constructed from online homework tracking reports and grades from an introductory economics class in the University of Pittsburgh. 255 students registered in this class and were requested to finish ten homework assignments on CourseWeb. This panel feature enables me to conduct both between and within subject analysis.

Each homework assignment consists of 20 multiple choice problems covering the material they learned during recent week. Each problem is 1 credit, therefore the full score is 20. Usually they have around one to two weeks to do the homework. CourseWeb will shut down the access to homework after it passes the deadline. A function on CourseWeb

called Statistics Tracking allows instructors to observe student homework activities. More specifically, the statistics tracking reports contain information on how many times a student opened each homework on any specific date (from here on, I call it an entry). I match these reports with students homework scores and exam scores to obtain a panel data set of every student's online homework activity and performance for each assignment.

### **3.3.2 Measurements of Procrastination, Time Cost and Performance**

Homework scores, normalized scores, and ranking in class are used as the performance measures. The higher the rank value is, at the higher percentile a student is in the class. Timing, or the degree of procrastination, is measured by how many days earlier than the deadline a student starts/submits/shows the maximum entries into the homework. Starting date is identified when the first entry happened, the submission date is identified when the last entry happened, and the doearly date is identified when the maximum entry happened. I refer to them as “startearly”, “submitearly”, and “doearly.”

There are two candidate measures for time cost. One is the number of lagged days between when a student started the homework and he/she submitted the homework, hereafter “lag.” This measure is very crude, because many students may decide to click into the homework and check it out, but did not intend to start it, which is supported by the summary statistics that students on average first enter into an assignment 9.944 days in advance of the deadline, but make the maximum entries and submit the homework 4.961 days and 2.839 days in advance. Therefore, I do not use lag as the time cost measure, but instead, use it as a control variable in some analysis, since students may behave/perform differently when given more days to work on a homework assignment.

The second candidate measure for time cost is how many times one entered into the homework page, hereafter “entrytimes.” It is not an accurate measure of how much time the student spent on the homework, but it is the best proxy to my ear. There could be students who only open the webpage and do nothing, but it is unlikely that one clicks the

Table 3.1: Summary Statistics

	(1)	(2)	(3)	(4)	(5)
VARIABLES	mean	sd	min	max	N
homework score	16.32	3.143	0	20	2,359
total entries	5.212	3.468	0	29	2,476
total days available	13.37	3.083	8	19	2,475
start early (days)	9.944	4.477	1	19	2,475
finish early (days)	2.839	2.183	1	19	2,475
do early (days)	4.961	4.117	1	19	2,475
normalized hw score	0	1	-8.540	2.049	2,359
normalized start early	0.723	0.240	0.0526	1	2,475
normalized finish early	0.225	0.178	0.0526	1	2,475
normalized do early	0.368	0.269	0.0526	1	2,475

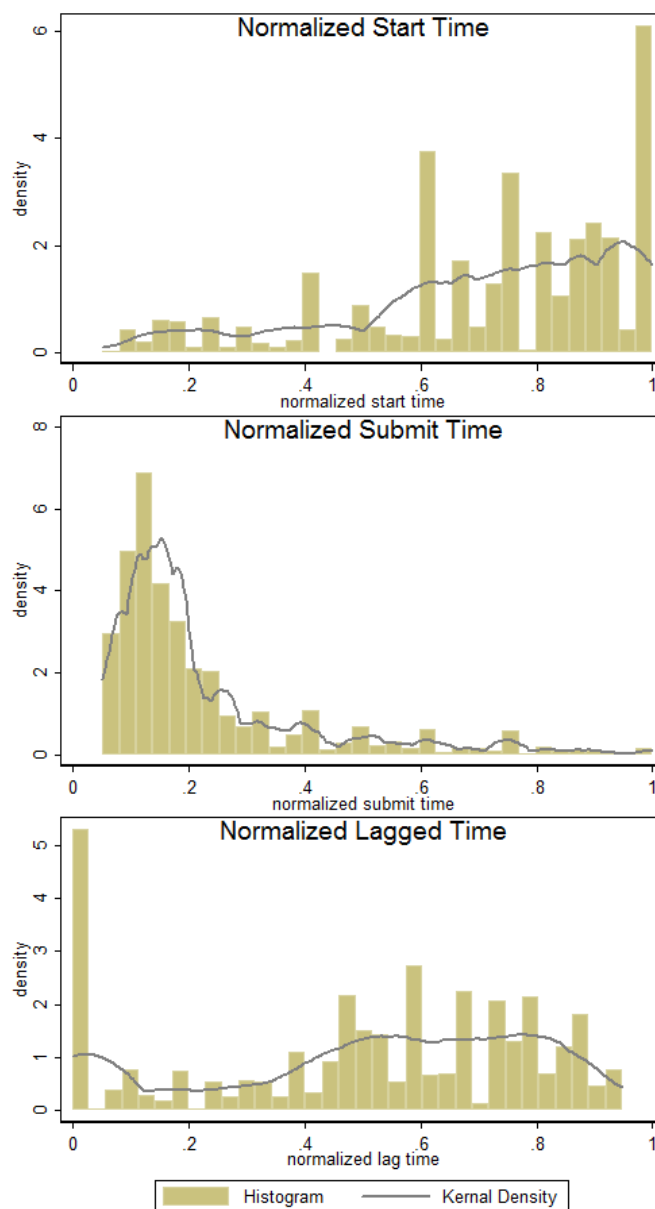
Note: The data set is constructed from online homework tracking logs for nine homework assignments in an introductory microeconomics course with 255 students, from different years in college and different majors. Each homework has 20 multiple choice questions, and the maximum score is 20. Total entry counts how many times a student click into the homework. Total days available means the number of days an online homework is open to students. Start/Finish/Do early is the number of days before the deadline when a student start/submit/has the maximum entries of the homework. Normalized start/finish/do early is constructed as the total days available divided by start/finish/do early.

webpage by chance all the time. The student must have the intention to do the homework or at least look at it. One reason why there are multiple entry times per day is that the webpage expires when it is inactive for a while. If a student opens the homework webpage and then turns to chat on Facebook for an hour, the webpage expires and he/she has to relog in the CourseWeb. In this sense, it is a fairly good measure of time cost on homework.

As for measures of starting and finishing doing the homework, I divide the number of days starting/finishing homework in advance of deadline by the number of days available. For any student on each homework assignment, the normalized early measure is a number from 0 to 1. The larger  $\text{norm startearly}/\text{norm doearly}/\text{norm submitearly}$  is, the earlier the student starts to look at/ works the most on/ submits the homework.

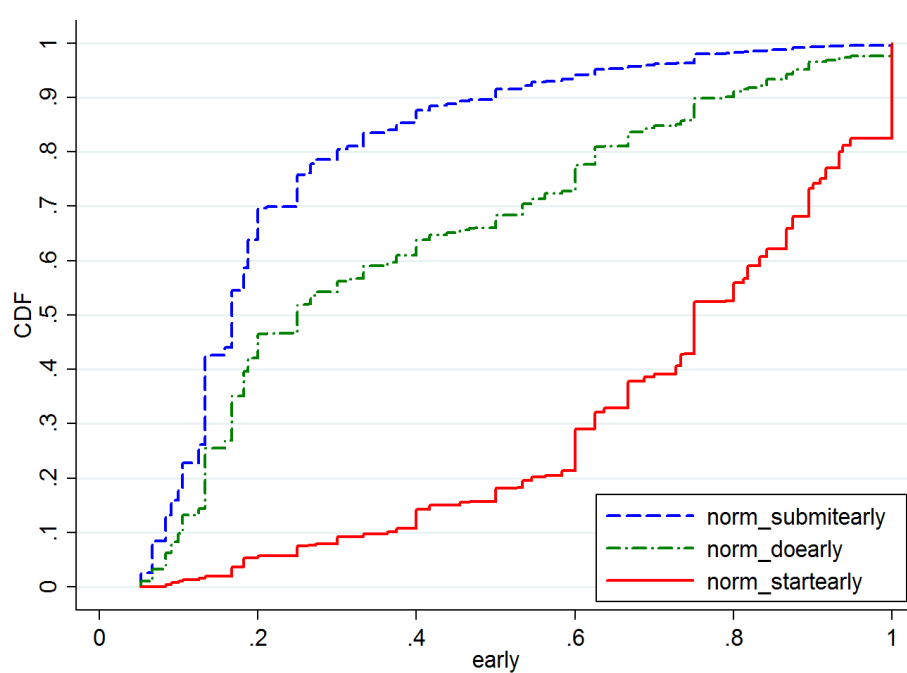
[Table 3.1](#) presents the summary statistics. Students usually submit the homework one to two days earlier than the deadline. [Figure 3.1](#) and [Figure 3.2](#) show us the general patterns on how early students start/submit/do their homework. A large portion of students check out homework early ( $\text{startearly}$  close to 1) and submit it late ( $\text{submitearly}$  close to 0).

Figure 3.1: Homework Start/Submit Timing Patterns



Note: In these density graphs, x-axis is the normalized measures of timing, with 0 being the deadline and 1 being the time when an assignment is posted.

Figure 3.2: Homework Start/Submit Timing Patterns



Note: In these density graphs, x-axis is the normalized measures of timing, with 0 being the deadline and 1 being the time when an assignment is posted.

### 3.4 EMPIRICAL METHODOLOGY AND RESULTS

Using online homework tracking data, I construct various performance measures, working early measures, and time cost measures. In this section, I conduct two sets of analysis to explore the relationship between time and work efficiency, measured by performance and time cost. Throughout the empirical analysis below, outcome variables are performance (score and rank) and time cost (total entries into the homework) and explanatory variables are work timing (absolute or normalized startearly/ doearly/ submitearly).

#### 3.4.1 Between-subject Analysis

In the between-subject analysis, I regress homework performance on procrastination measures and its quadratic form. This exercise shows us whether those students who start or finish early do better than those who start or finish late, and whether the relationship is non-linear.

$$Y_{ij} = \alpha_0 + \alpha_1 E_{ij} + \alpha_2 E_{ij}^2 + \mu_{ij} \quad (3.2)$$

where  $Y_{ij}$  are performance measures and time cost measures;  $E_{ij}$  are procrastination measures (how early a student  $i$  starts/does/submits his/her homework  $j$ ).

Table 3.2 shows the results of Equation 3.2 when independent variables are homework score, normalized homework score (with mean 0 and standard deviation 1), homework score rank in class (larger value means higher ranking) and dependent variables are measures of how early a student starts and submits an assignment. Normalized start/finish early is constructed as the total days available to do the homework divided by the number of days before the deadline when a student start or finish the homework. For example, if a student starts 2 days before the deadline, submits it 1 day before the deadline and the total days available is 8 days, then norm\_startearly equals to  $2/8=0.25$  and norm\_finishearly equals to  $1/8=0.125$ . Panels A/ B look at the relationship between performance and start early/

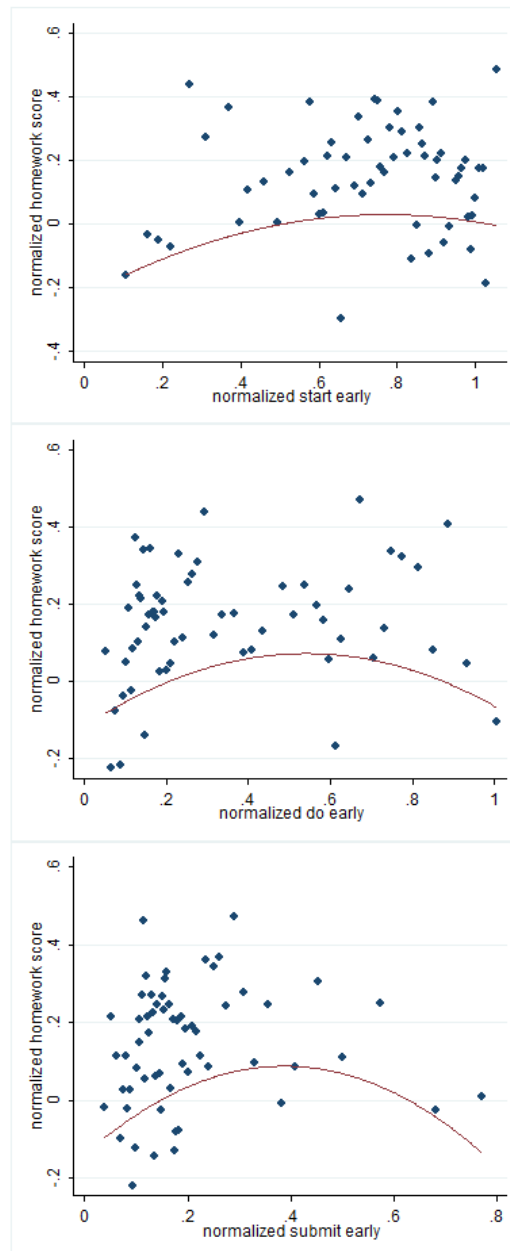


finish early. Results for do early are similar with those for finish early. For each panel, first three columns show the result of quadratic effect of timing on performance. The following three columns add control of total entries into the homework.

[Table 3.2](#) and [Table 3.3](#) show that students who start early generally do better than the others, but also spend more time on the homework than the others. As for people who finish early, they not only do better than the others, but also spend less time on it. Both timing-performance relationship and timing-time cost relationship are in a nonlinear fashion, as shown in a binned scatter plot in [Figure 3.3](#) and [Figure 3.4](#).

Figure 3.3: Between-Subject Analysis:

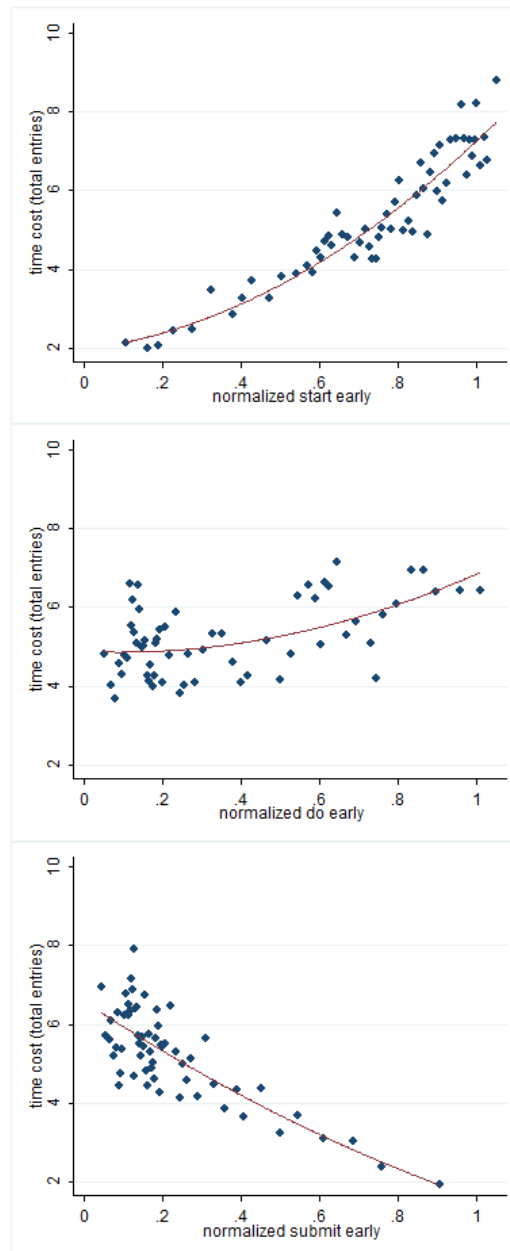
Relationship between Start/Do/Submit Early and Homework Performance



Note: These are binned scatter plots with quadratic fitted lines, controlling for midterm and final scores.

Figure 3.4: Between-Subject Analysis:

Relationship between Start/Do/Submit Early and Time Cost



Note: These are binned scatter plots with quadratic fitted lines, controlling for midterm and final scores.

Table 3.2: Between-subject analysis: Performance and Timing

Panel A. performance and start early						
Depend. Var.	(1)	(2)	(3)	(4)	(5)	(6)
	homework score	normalized hw score	score ranking	homework score	normalized hw score	score ranking
norm_startearly	3.107** (1.211)	1.110*** (0.393)	85.32*** (23.84)	2.990** (1.209)	1.066*** (0.391)	81.95*** (23.78)
norm_startearly sq	-1.086 (0.975)	-0.666** (0.315)	-53.85*** (19.57)	-1.431 (0.973)	-0.795** (0.313)	-63.83*** (19.35)
hw entry times				0.0838*** (0.0249)	0.0314*** (0.00792)	2.421*** (0.613)
Observations	2,358	2,358	2,358	2,358	2,358	2,358
R-squared	0.019	0.007	0.008	0.026	0.016	0.022
Panel B. performance and submit early						
Depend. Var.	(7)	(8)	(9)	(10)	(11)	(12)
	homework score	normalized hw score	score ranking	homework score	normalized hw score	score ranking
norm_finishearly	6.125*** (1.501)	2.327*** (0.478)	163.2*** (30.23)	7.101*** (1.503)	2.583*** (0.482)	181.2*** (30.19)
norm_finishearly sq	-5.318*** (1.837)	-2.622*** (0.556)	-198.7*** (35.50)	-5.616*** (1.822)	-2.700*** (0.551)	-204.2*** (35.02)
hw entry times				0.150*** (0.0233)	0.0392*** (0.00744)	2.773*** (0.563)
Observations	2,358	2,358	2,358	2,358	2,358	2,358
R-squared	0.015	0.013	0.014	0.041	0.030	0.037

Note: Standard errors in parentheses, \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . Independent variables are homework score, normalized homework score (with mean 0 and standard deviation 1), homework score rank in class (bigger number means higher ranking). Dependent variables are measures of how early a student starts and submits an assignment. Normalized start/finish early is constructed as the total days available to do the homework divided by the number of days before the deadline when a student start or finish the homework.

### 3.4.2 Within-subject Analysis

Although the between-subject analysis strongly supports the non-linear predictions by efficient procrastination framework, it could be explained by selection. Students who start to work early may be those who are struggling in class. Therefore, even though they work early, they still get a low score in homework assignments and spend more time than other students. This alternative explanation may confound the inferences on the efficient procrastination.

To rule out the alternative explanation from the self-selection, I take advantage of the panel feature of the data and add individual fixed effects to conduct within-subject analysis. I use homework score, normalized homework score, and ranking in class as the performance measures and normalized starting early, normalized submitting early, and normalized do early as the main explanatory variables. The larger the number for rank is, higher the student ranks in the class and better he/she performs. Normalized measures of starting early and finishing early are generated by dividing startearly and finishearly by the available time for each homework.

$$Y_{ij} = \beta_i + \beta_1 E_{ij} + \beta_2 E_{ij}^2 + \epsilon_{ij} \quad (3.3)$$

where  $Y_{ij}$  are performance measures and time cost measures;  $\beta_i$  is a vector of individual fixed effects,  $E_{ij}$  are procrastination measures (how early a student  $i$  starts/does/submits his/her homework  $j$ ).

Table 3.4 presents results of fixed-effects OLS regressions using data on 255 students' 10 homework activities in an introductory economics class. 142 observations are dropped because scores are zero or missing. Independent variables are homework score, normalized homework score (with mean 0 and standard deviation 1), homework score rank in class (bigger number means higher ranking). Dependent variables are measures of how early a student starts and submits an assignment. Normalized start/finish early is constructed as the total days available to do the homework divided by the number of days before the

Table 3.3: Between-subject analysis: Time cost (total entries) and timing

Panel A: Absolute Early Measures as Explanatory Variables						
Dependent Variable: Total Entries into Homework						
	(1)	(2)	(3)	(4)	(5)	(6)
Early Measure	startearly	startearly	finishearly	finishearly	doearly	doearly
early	1.074 (1.141)	0.547 (1.134)	-6.700*** (1.379)	-1.221 (1.397)	-0.285 (1.153)	3.933*** (1.149)
early squared	4.192*** (1.109)	3.801*** (1.109)	1.665 (1.427)	-3.009** (1.460)	2.215* (1.293)	-2.461* (1.281)
total days available		0.212*** (0.0188)		0.309*** (0.0191)		0.373*** (0.0187)
Observations	2,475	2,475	2,475	2,475	2,475	2,475
R-squared	0.195	0.225	0.076	0.140	0.021	0.124
Number of stu	255	255	255	255	255	255
Panel B: Normalized Early Measures as Explanatory Variables						
Dependent Variable: Total Entries into Homework						
	(7)	(8)	(9)	(10)	(11)	(12)
Early Measure	normalized startearly	normalized startearly	normalized finishearly	normalized finishearly	normalized doearly	normalized doearly
early	0.494*** (0.0433)	0.497*** (0.0435)	-0.297*** (0.0812)	-0.105 (0.0781)	0.187*** (0.0602)	0.284*** (0.0580)
early squared	-0.00656*** (0.00236)	-0.00407 (0.00268)	0.000567 (0.00549)	-0.0166*** (0.00533)	0.000546 (0.00417)	-0.0114*** (0.00414)
total days available		-0.0961*** (0.0358)		0.382*** (0.0186)		0.332*** (0.0186)
Observations	2,475	2,475	2,475	2,475	2,475	2,475
R-squared	0.224	0.227	0.033	0.146	0.054	0.127
Number of stu	255	255	255	255	255	255

Note: Absolute early measures are the number of days before deadlines; normalized early measures are absolute ones divided by total days available.

deadline when a student start or finish the homework. For example, if a student starts 2 days before the deadline, submits it 1 day before the deadline and the total days available is 8 days, then  $\text{norm\_startearly}$  equals to  $2/8=0.25$  and  $\text{norm\_finishearly}$  equals to  $1/8=0.125$ . Panels A and B separately look at effect of start early and finish early. For each panel, first three columns show the result of quadratic effect of timing on performance.

I plot binned scatter plot and predict a fitted quadratic line for relationship between performance and timing, controlling for student individual fixed effects. In [Figure 3.5](#), the shape of fitted quadratic line in normalized finish early is the closest to the theoretical predictions: hump shape curves of performance and working early. Those for normalized do early and normalized start early is not clear enough. This indicates that, for each individual student, when he/she submits the homework not too early nor too late, he/she gets a higher homework score.

Similarly, [Table 3.5](#) and [Figure 3.6](#) show the relationship between time cost and timing, controlling for student fixed effects. The relationship between start/finish early and time cost is close to linear: when students start early, they enter into the homework website more times, which indicates that they spend more time on the homework; when students finish early, they submit homework early and enters into the homework website less times. The relationship between do early and time cost matches the best with the prediction from the conceptual model in [section 3.2](#): a U-shaped curve. This indicates that, for each individual student, when he/she focuses to do the homework not too early nor too late, he/she spends the least time on this assignment.

Table 3.4: Within-Subject Analysis: Normalized Timing and Homework Performance

Panel A. effect of start early						
Depend. Var.	(1) homework score	(2) normalized hw score	(3) score ranking	(4) homework score	(5) normalized hw score	(6) score ranking
norm_startearly	1.853 (1.146)	0.693* (0.354)	62.75*** (22.03)	1.725 (1.146)	0.645* (0.354)	59.50*** (22.00)
norm_startearly sq	-0.361 (0.907)	-0.469* (0.280)	-44.99*** (17.43)	-0.512 (0.908)	-0.526* (0.280)	-48.82*** (17.43)
hw entry times				0.0570** (0.0229)	0.0214*** (0.00707)	1.448*** (0.439)
Student Fixed Effect	Y	Y	Y	Y	Y	Y
Observations	2,358	2,358	2,358	2,358	2,358	2,358
R-squared	0.015	0.002	0.004	0.017	0.007	0.009
Number of stu	255	255	255	255	255	255
Panel B. effect of finish early						
Depend. Var.	(7) homework score	(8) normalized hw score	(9) score ranking	(10) homework score	(11) normalized hw score	(12) score ranking
norm_finishearly	1.083 (1.333)	0.603 (0.410)	65.03** (25.46)	2.209* (1.337)	0.814** (0.413)	77.86*** (25.68)
norm_finishearly sq	0.218 (1.676)	-0.811 (0.515)	-101.2*** (32.02)	-0.357 (1.666)	-0.919* (0.515)	-107.7*** (31.99)
hw entry times				0.122*** (0.0213)	0.0229*** (0.00658)	1.389*** (0.409)
Student Fixed Effect	Y	Y	Y	Y	Y	Y
Observations	2,358	2,358	2,358	2,358	2,358	2,358
R-squared	0.004	0.001	0.006	0.019	0.007	0.011
Number of stu	255	255	255	255	255	255

Note: Standard errors in parentheses, \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . This table presents results of fixed-effect OLS regressions using data on 255 students' 10 homework activities in an introductory economics class. 142 observations are dropped because scores are zero or missing. Independent variables are homework score, normalized homework score (with mean 0 and standard deviation 1), homework score rank in class (bigger number means higher ranking). Normalized start/finish early is constructed as the total days available to do the homework divided by the number of days before the deadline when a student start or finish the homework. Panel A and B separately look at effect of start early and finish early. For each panel, first three columns show the result of quadratic effect of timing on performance. The following three columns add control variable "total entries into an assignment".

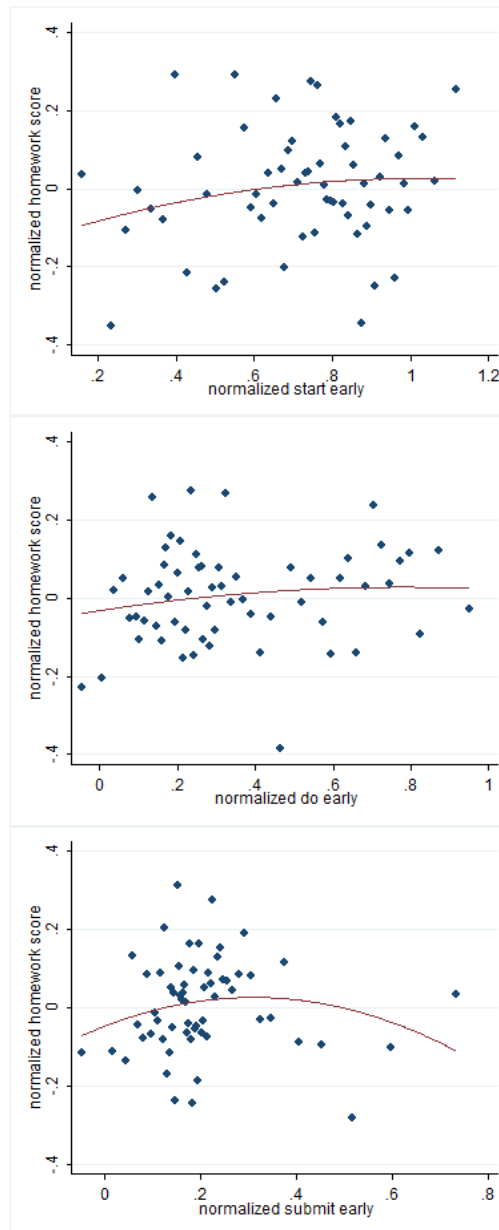


Table 3.5: Within-Subject Analysis: Timing and Time Cost (total entries)

Panel A: Absolute Early Measures as Explanatory Variables						
Dependent Variable: Total Entries into Homework						
Early Measure	(1)	(2)	(3)	(4)	(5)	(6)
startearly	startearly	startearly	finishearly	finishearly	doearly	doearly
early	0.403*** (0.0466)	0.396*** (0.0467)	-0.303*** (0.0752)	-0.0660 (0.0683)	0.179*** (0.0496)	0.282*** (0.0461)
early squared	-0.00400* (0.00227)	-0.00513** (0.00235)	0.00640 (0.00562)	-0.0139*** (0.00512)	0.000265 (0.00311)	-0.0130*** (0.00295)
total days available		0.0499* (0.0282)		0.387*** (0.0167)		0.356*** (0.0182)
Student FE	Y	Y	Y	Y	Y	Y
Observations	2,475	2,475	2,475	2,475	2,475	2,475
R-squared	0.266	0.267	0.026	0.216	0.069	0.206
Number of stu	255	255	255	255	255	255
Panel B: Normalized Early Measures as Explanatory Variables						
Dependent Variable: Total Entries into Homework						
Early Measure	(7)	(8)	(9)	(10)	(11)	(12)
normalized startearly	normalized startearly	normalized startearly	normalized finishearly	normalized finishearly	normalized doearly	normalized doearly
early	2.070* (1.073)	1.227 (1.023)	-9.232*** (1.171)	-1.841 (1.171)	-1.850* (0.954)	2.791*** (0.889)
early squared	2.589*** (0.847)	2.006** (0.807)	5.064*** (1.376)	-1.128 (1.332)	3.574*** (0.984)	-1.570* (0.920)
total days available		0.266*** (0.0175)		0.328*** (0.0184)		0.379*** (0.0172)
Student FE	Y	Y	Y	Y	Y	Y
Observations	2,475	2,475	2,475	2,475	2,475	2,475
R-squared	0.185	0.262	0.095	0.209	0.025	0.201
Number of stu	255	255	255	255	255	255

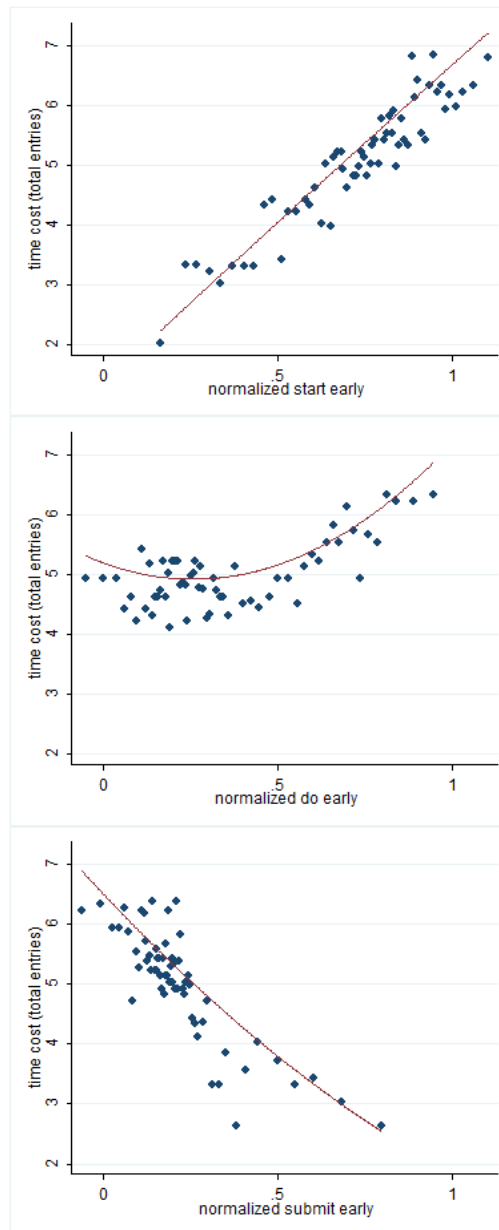
Note: Absolute early measures are the number of days before deadlines; normalized early measures are absolute ones divided by total days available.

Figure 3.5: Within-Subject Analysis:  
Relationship between Start/Do/Submit Early and Homework Performance



Note: These are binned scatter plots with quadratic fitted lines, controlling for individual fixed effects.

Figure 3.6: Within-Subject Analysis:  
Relationship between Start/Do/Submit Early and Time Cost



Note: These are binned scatter plots with quadratic fitted lines, controlling for individual fixed effects.

### 3.5 CONCLUSION AND FUTURE WORK

This paper provides a new perspective showing that procrastination is not necessarily counter-productive. Given a task with a deadline, appropriate postponing can raise the level of focus, therefore not only improve the performance but also shorten the time spent on the task. Analysis of a data set of college students' procrastination on homework supports the hypothesis. It shows that working early affects students' efficiency, measured by homework performance and time cost, in a nonlinear way.

Starting to work too early or too late brings under or over motivation and therefore lowers one's efficiency, which leads to low performance and high time cost. It supports the efficient procrastination framework that people may benefit from appropriate amount of procrastination because of efficiency gain (better performance with less time cost).

One limitation of this study is that measures for the timing when students do their homework are coarse. Normalized start/do/finish early are the date they first look at the homework, the maximum entries into the homework, and the submission date of the homework. Although do/finish early can serve as proxies of when students did their homework, we still do not know exactly when it happened. Further accuracy of the timing could possibly be obtained with more information from the online tracking system.

There are possible limitations of this study. In the within-subject analysis, if a student starts/does/submits an assignment early because he/she feels unfamiliar with the material that week, the time cost may be high and the performance may still be low. In order to test the validity of this concern, we need information on the prior knowledge of a student to the content of each homework. A quiz or a survey at the beginning of a semester would provide us with such information.

Another confound is if a student starts/does/submits a homework assignment early because he/she is busy that week and has to work that way to get around the schedule. An easy way to obtain information to test this concern is to add an additional question

into each homework about how busy they are this week.

Granted, it is hard to find the right balance of procrastination and calculating the optimal timing of work. So when is not too early nor too late? This study cannot answer this question. Possible intervention and study to improve learning and work efficiency and to find out the optimal timing calculation algorithm would be sending text messages or other types of reminders to nudge people to work at different times, then evaluate their efficiency, performance, and time costs.

## APPENDIX

### A.1 APPENDIX FOR CHAPTER 1

The original data sets for 2005-2008 and 2006-2009 cohorts were in two pieces and for 2007-2010 and 2008-2011 cohorts were in three pieces, with the information and number of observations listed in the following table. Note that lottery data set B and junior high graduation data set C contain information of all students in the city, while primary school scores in data set A for 2007-2010 cohort is only for one district and 2008-2011 cohort for two districts, one of which is the same as 2007-2010 cohort.

For the composition analysis, I merge data set A and B for 2007-2010 and 2008-2011 cohort for one shared district. Since there is not much time lag between when A and B were collected, i.e. when students graduated from elementary school, the matching rates are high. I only drop few duplicates and the matching rate is around 95%. The unmatched may move to another city or because of mis-typed names that cannot identify by pronunciation of names.

For the instrumental QTE analysis, I merge data set B and C for three cohorts using name, gender, birth date and middle school, the linkage rate is lower. Possible reasons include mis-typed names, incomplete information on birth date, noncompliance of the lottery assignment and transfer. In order to link as many students' record as possible, I

gradually relax the criteria of matching.

After each stage, I take out the matched observations and use the remaining unmatched observations in both data sets to do the next stage of matching. Stage 1 gives us the most reliable matches. Stage 2 captures people whose names were mistyped. Stage 3 captures people whose birth date information is inaccurate. Stage 4 and 5 captures students transferred to another middle school. <sup>1</sup> In total, the matching rate is around 80%.

---

<sup>1</sup>We would expect that the transferred middle school should be, on average, of higher quality than the original one. There are of course other reasons causing transfer, such as moving and transferring to school closer to home.

Table A1: Description of Data Sets

Cohort 2007-2010	
Data Set A: One district 6th	Name, elementary school, 6th grade Chinese, math scores grade scores
Data Set B: City lottery record	Name, elementary school, gender, birth date, class, admission channel, lottery choice and outcome, middle school admitted, parents' political status, hukou, parents' occupation, address, hometown, ethnicity, political status
Data Set C: City 9th grade record	Name, gender, birth date, middle school attended, middle school graduation score, high school admitted.
Cohort 2006-2009	
Data Set B: City lottery record	Name, elementary school, gender, birth date, class, admission channel, lottery choice and outcome, middle school admitted, hukou, address, hometown, ethnicity
Data Set C: City 9th grade record	Name, gender, birth date, middle school attended, middle school graduation score, high school admitted.
Cohort 2005-2008	
Data Set B: City lottery record	Name, elementary school, gender, admission channel, lottery choice and outcome, middle school admitted
Data Set C: City 9th grade record	Name, gender, birth date, middle school attended, middle school graduation score.



The following three tables present us with the descriptive statistics of the data for three cohorts: 2005-2008, 2006-2009 and 2007-2010. I divide all students into three groups: pre-admitted, noncompetitive lottery takers and competitive lottery takers. As we can see, pre-admitted students have higher middle school graduation scores and better family background (in terms of father and mother political status and hukou possession <sup>2</sup>, attend better schools and get higher scores in junior high graduation exams.

Table A2: Individual Level Data: Summary Statistics 2005

Sample	(1) All	(2) policy	(3) preadmission	(4) lottery	(5) lottery&policy	(6) lottery&nonpolicy
9th grade score	22.39	20.79	24.61	20.58	19.80	20.72
Non-academic Evaluation	13.20	12.89	13.80	12.77	12.83	12.76
Higher than 90 percentile	0.442	0.0582	0.464	0.388	0.0531	0.0632
Normalized 9th grade score	0.746	0.693	0.820	0.686	0.660	0.691
Normalized Non-academic Evaluation	0.825	0.806	0.862	0.798	0.802	0.798
female	0.475	0.469	0.477	0.457	0.463	0.456
transfer	0.200	0.220	0.0730	0.371	0.369	0.372
policy	0.0994	1	0.00608	0.155	1	0
preadmission	0.385	0.0264	1	0	0	0
lottery	0.240	0.365	0	1	1	1
winlottery	0.480	0.399		0.480	0.399	0.494
obs	12,964	1,289	7,467	4,653	471	2,563

Note: Column 2 describes policy school students, column 3 describes students who were pre-admitted, column 4 describes students who chose an over-subscribed school and assigned by lottery; column 5 describes students who were assigned by lottery to a policy school; column 6 describes students who were assigned by lottery to a non-policy school. Non-academic evaluation is consist of teacher and self-rated measures of four abilities, including civics, learning ability, atheistic ability, and practical ability.

<sup>2</sup>Hukou equals to 1 if the student has the residency record of Changsha. In China, residency record is very important because it gives you access to many benefits in the city, including health care, pension insurance and employment advantages.

Table A3: Individual Level Data: Summary Statistics 2006

Sample	(1) All	(2) policy	(3) preadmission	(4) lottery	(5) lottery&policy	(6) lottery&nonpolicy
9th grade score	22.67	21.68	24.57	21.75	21.90	21.72
Normalized 9th grade score	0.756	0.723	0.819	0.725	0.730	0.724
Academic high school	0.837	0.760	0.951	0.785	0.786	0.785
Elite high school	0.369	0.295	0.575	0.224	0.288	0.208
Non-academic Evaluation	14.59	14.11	15.11	14.37	14.22	14.41
Normalized Non-academic Evaluation	0.912	0.882	0.945	0.898	0.888	0.901
Imputed 9th grade score	22.72	22.06	24.57	21.85	22.23	21.75
female	0.469	0.456	0.489	0.460	0.461	0.460
hukou	0.778	0.754	0.862	0.776	0.752	0.784
missing hukou	0.174	0.0355	0.233	0.103	0.0383	0.121
preadmission	0.387	0.000480	1	0	0	0
policy	0.125	1	0.000155	0.220	1	0
winlottery	0.546	0.197		0.546	0.197	0.645
transfer	0.0910	0.101	0.0881	0.0852	0.120	0.0755
obs	16,665	2,082	6,446	5,694	1,254	4,440

Note: Column 2 describes policy school students, column 3 describes students who were pre-admitted, column 4 describes students who chose an over-subscribed school and assigned by lottery; column 5 describes students who were assigned by lottery to a policy school; column 6 describes students who were assigned by lottery to a non-policy school. % academic high school indicates the percentage of 9th grade graduates attending an academic high school; some other graduates attend vocational schools or stop schooling. Non-academic evaluation is consist of teacher and self-rated measures of four abilities, including civics, learning ability, atheistic ability, and practical ability. Imputed 9th grade score is constructed by assigning the highest grade of their cohort to the missing grade of direct admitted students who did not take the exam. Having city hukou means that a student is born in city and enjoys the public goods of that city; it is often used as a measure of socioeconomic background.

Table A4: Falsification Test for Treatment Effects

	(1)	(2)	(3)	(4)	(5)	(6)
Depend.Variable	Normalized 9th grade score	Normalized school ranking			% elite high school	
Panel A. All Schools						
postXlow-performing	0.0734 (0.0957)	-0.0311 (0.0873)	0.0883*** (0.0309)	0.0393 (0.0282)	0.0185 (0.0161)	0.0171 (0.0151)
Observations	441	441	441	441	441	441
R-squared	0.062	0.006	0.071	0.010	0.130	0.027
Number of schools	80	80	80	80	80	80
Panel B. Balanced sample						
postXlow-performing	-0.0183 (0.156)	-0.156 (0.145)	0.0540 (0.0501)	-0.00873 (0.0465)	-0.0101 (0.0269)	-0.0113 (0.0255)
Observations	160	160	160	160	160	160
R-squared	0.158	0.022	0.143	0.006	0.217	0.056
Number of schools	20	20	20	20	20	20
School FE	Y	Y	Y	Y	Y	Y
Year FE	Y	N	Y	N	Y	N
Time Trend	N	Y	N	Y	N	Y

This table reports falsification check for treatment effects. Instead of estimating treatment effects on policy schools reported in [Table 1.4](#), it tests whether the post-2007 improvement holds true for any low-performing school. Each cell reports the coefficient of “postXlow-performing”, which equals to 1 for low-performing schools (defined by below average in 2004) after 2007. The sample ranges from 2004 to 2011 and drops all the policy schools. Each column has dependent variable listed on top row and control variables indicated at the bottom three rows. Panel B takes schools with observed performance in all eight years. Standard errors in parentheses. Significance level indicated by \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

## A.2 APPENDIX FOR CHAPTER 2

Table A5: Balance Check for the Whole Sample

		Control		Treatment		Stat. Sig. Different?
		mean	s.d.	Mean	s.d.	
<b>Class size</b>	Average class size	36.5	4.48	40.5	2.66	Y
<b>Personal Characteristics</b>	Female percentage	0.47	0.5	0.51	0.5	
	Baseline Std Math					
	Grade 7	-0.08	0.9	0.13	0.91	Y
	Grade 8	-0.24	1.13	0.13	1.03	Y
	Baseline MHT	38.09	14.37	39.46	13.52	
<b>Family Characteristics</b>	Father's education	2.14	1.04	2.12	1.03	
	Mother's education	1.75	0.96	1.8	0.97	
	Family size	4.76	1.12	4.81	1.2	
	Boarding roommate	6.01	3.79	5.54	4.47	
<b>Study attitude</b>	like school (1-10)	7.639	2.805	7.36	2.691	
	class participation	5.806	6.299	4.286	5.373	
	ask questions	2.426	0.541	2.459	0.532	
	self read	2.366	0.595	2.244	0.534	
	extra exercise	2.106	0.634	2.074	0.485	
	Class leader	1.347	0.477	1.384	0.487	
<b>MHT subcategory scores</b>	Learning stress	8.81	2.99	8.44	3.12	
	Social stress	3.92	2.35	4.36	2.3	
	Anti-social	2.82	2.18	2.84	2.09	
	Self-guilty	5.43	2.51	5.43	2.39	
	Over-sensitive	5.4	2.2	5.85	2.06	
	Physical-signs	5.49	2.89	5.46	2.81	
	fear	3.98	2.73	3.96	2.74	
	impulsiveness	3.1	2.47	3.24	2.19	
<b>Non-academic Behavior</b>	bullied	1.574	0.724	1.541	0.657	
	argument	2.741	4.862	1.855	2.708	Y
	fight	1.065	2.496	0.694	1.413	Y
	Late hw	1.597	0.905	1.707	0.956	
	Late for class	1.458	0.681	1.426	0.615	
	absence	1.727	0.712	1.777	0.778	

Note: First three columns takes the lower half performing students based on baseline math score in the control classes as the control group for mentees. Similarly for mentors. "Control" and "Treatment" column show the means and standard deviations in parentheses. "Stat. sig. diff?" column indicates whether the control and treatment group has statistically significantly different characteristics.

Table A6: Balance Check for Mentor/Mentee Subsamples

	Mentee			Mentor		
	Control	Treatment	Diff	Control	Treatment	Diff
family size	4.718 (1.089)	4.726 (1.129)	0.00786 (0.143)	4.818 (1.155)	4.907 (1.268)	0.0886 (0.165)
boarding length (year)	2.479 (2.286)	2.507 (2.120)	0.0279 (0.285)	2.328 (2.129)	2.156 (1.974)	-0.172 (0.281)
# boarding roommates	6.393 (3.893)	5.702 (4.277)	-0.692 (0.526)	5.556 (3.643)	5.377 (4.684)	-0.178 (0.566)
class participation	5.017 (4.657)	3.605 (2.686)	<b>-1.412***</b> (0.494)	6.737 (7.731)	5.009 (7.146)	<b>-1.729*</b> (1.020)
ask question	2.385 (0.539)	2.484 (0.518)	0.0993 (0.0681)	2.475 (0.541)	2.432 (0.547)	-0.0425 (0.0741)
read	2.342 (0.604)	2.266 (0.543)	-0.0758 (0.0741)	2.394 (0.586)	2.220 (0.525)	<b>-0.174**</b> (0.0762)
extra exercise	2.103 (0.662)	2 (0.460)	-0.103 (0.0738)	2.111 (0.604)	2.153 (0.500)	0.0414 (0.0762)
bullied	1.615 (0.808)	1.669 (0.707)	0.0540 (0.0980)	1.525 (0.612)	1.407 (0.573)	-0.118 (0.0810)
argument	3.325 (6.270)	1.847 (2.157)	<b>-1.478**</b> (0.611)	2.051 (2.106)	1.864 (3.197)	-0.186 (0.363)
fight	1.302 (2.964)	0.952 (1.701)	-0.350 (0.315)	0.788 (1.774)	0.424 (0.964)	<b>-0.364*</b> (0.199)
late homework	1.654 (3.090)	1.875 (3.001)	0.221 (0.336)	1.167 (1.786)	1.538 (2.375)	0.371 (0.290)
late class	0.987 (1.822)	0.875 (1.561)	-0.112 (0.210)	1.071 (1.736)	0.996 (1.353)	-0.0749 (0.195)
like school (1-10)	7.829 (2.922)	7.145 (3.017)	<b>-0.684*</b> (0.383)	7.414 (2.657)	7.585 (2.292)	0.171 (0.340)

Note: First three columns takes the lower half performing students based on baseline math score in the control classes as the control group for mentees. Similarly for mentors. “Control” and “Treatment” column show the means and standard deviations in parentheses. “Diff” column shows the difference and standard errors in parentheses, with significance level indicated by \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table A7: Main Results using the Pooled Sample

Regression Model	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) Probit	(6) Probit
Dependent Variable	Post std math	Diff std math	Post mental health	Diff mental health	Dropout	Dropout_any
<b>treat</b>	0.191 (0.154)	0.0511 (0.135)	-1.611 (1.363)	-1.397 (1.432)	-0.456 (0.294)	-0.386 (0.285)
grade8	-0.0587 (0.180)	-0.0245 (0.102)	-2.064 (1.452)	-2.207 (1.537)	0.253 (0.280)	0.320 (0.287)
female	-0.0547 (0.112)	-0.0260 (0.129)	0.104 (0.740)	1.421 (1.177)	-0.164 (0.245)	-0.157 (0.144)
Base std math	0.457*** (0.000)					
Base mental health			0.766*** (0.000)			
Observations	433	433	390	390	457	457
R-squared	0.22	0.001	0.536	0.023		

Significance level \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . Columns (1)-(4) are OLS regression results with robust standard errors corrected for small number of clusters using wild cluster bootstrapping with 1000 replications in parentheses. (5)-(6) are probit regression results with robust standard errors clustered at class level in parentheses.

Standardized Math Score has a mean of 0 and standard deviation of 1 for each grade in each exam. Mental health score is calculated according to a 100 question survey, with higher score being better mental health condition. Dropout in column (5) equals to 1 if we know a student dropped out of school. The dependent variable dropout\_any in column (6) equals to 1 if a student dropped out or if he/she is missing from our sample without knowing the reason.

Table A8: Heterogeneous Treatment Effect for Different Grade and Gender

<b>Panel A: Mentee and Placebo Mentee Subsample</b>				
	(1)	(2)	(3)	(4)
VARIABLES	post std	math score	post mental health	
treat	-0.181 (0.175)	-0.0374 (0.0802)	-3.423* (1.795)	-4.867 (3.251)
female	-0.0130 (0.162)	0.0209 (0.117)	0.613 (1.366)	0.801 (0.966)
grade8	-0.0455 (0.123)	0.0654 (0.197)	-1.693 (1.530)	-3.483 (2.422)
treat X female	0.0695 (0.233)		0.560 (2.404)	
treat X grade 8		-0.221 (0.318)		3.523 (2.510)
Baseline Control	Y	Y	Y	Y
Observations	224	224	200	200
R-squared	0.079	0.083	0.499	0.501
<b>Panel B: Mentor and Placebo Mentor Subsample</b>				
	(1)	(2)	(3)	(4)
VARIABLES	post std	math score	post mental health	
treat	0.451 (0.309)	0.239 (0.234)	1.272 (1.452)	0.120 (4.340)
female	-0.273 (0.237)	-0.140 (0.138)	0.901 (2.057)	-0.691 (1.640)
grade8	-0.112 (0.263)	-0.460 (0.616)	-2.694 (1.873)	-2.451 (1.561)
treat X female	0.248 (0.284)		-2.897 (2.730)	
treat X grade 8		0.654 (0.530)		-0.496 (4.585)
Baseline Control	Y	Y	Y	Y
Observations	209	209	190	190
R-squared	0.173	0.195	0.593	0.59

Robust standard errors in parentheses, corrected for small number of clusters using wild cluster bootstrapping with 1000 replications. Significance level indicated by \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . Panel A takes the subsamples of mentees in the treatment group and the lower half performing students (based on baseline math score) in the control group. Similarly for Panel B. Each regression includes baseline control of the outcome variables.

Table A9: Student Suggestions

Category	Detail	# stu. response
Organization and arrangement		57
	more time/frequency	20
	more contact	10
	fix time of the day to do peer mentoring	8
	seat together	7
	more monitoring	5
	more friend conversations	3
	separate study & non-study students	2
	fix subject for each day	2
Matching mechanism		46
	More ppl in a group (esp. 4 ppl group)	22
	Self-choice pair	10
	Top-down pair	8
	Top-top pair	3
	Mix gender group	3
Incentive methods		18
	More prize for academic improvement/excellence	11
	No snack or change snack to other prizes (pen, book)	7
No Change/ No Suggestion		53



## BIBLIOGRAPHY

- Abadie, Alberto, Joshua Angrist, and Guido Imbens**, “Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings,” *Econometrica*, 2002, 70 (1), 91–117.
- Abdulkadiroğlu, Atila, Joshua D Angrist, Susan M Dynarski, Thomas J Kane, and Parag A Pathak**, “Accountability and flexibility in public schools: Evidence from Boston’s charters and pilots,” *The Quarterly Journal of Economics*, 2011, 126 (2), 699–748.
- Ajayi, Kehinde F**, “School Choice and Educational Mobility: Lessons from Secondary School Applications in Ghana,” *Working Paper*, 2011.
- Akerlof, George A**, “Procrastination and obedience,” *The American Economic Review*, 1991, 81 (2), 1–19.
- Andrabi, Tahir, Jishnu Das, and Asim Ijaz Khwaja**, “Report cards: The impact of providing school and child test scores on educational markets,” *Unpublished Working Paper*, 2009.
- Angrist, Joshua D and Jörn-Steffen Pischke**, *Mostly harmless econometrics: An empiricist’s companion*, Princeton university press, 2008.
- , **Daniel Lang, and Philip Oreopoulos**, “Incentives and services for college achievement: Evidence from a randomized trial,” *American Economic Journal: Applied Economics*, 2009, pp. 136–163.
- , **Eric P Bettinger, Erik Bloom, Elizabeth King, and Michael Kremer**, “Vouchers for private schooling in Colombia: Evidence from a randomized natural experiment,” *American Economic Review*, 2002, pp. 1535–1558.
- , **Parag A Pathak, and Christopher R Walters**, “Explaining Charter School Effectiveness,” *American Economic Journal: Applied Economics*, 2013, 5 (4), 1–27.

- , **Sarah R Cohodes, Susan M Dynarski, Parag A Pathak, and Christopher R Walters**, “Stand and Deliver: Effects of Boston’s Charter High Schools on College Preparation, Entry, and Choice,” *NBER Working Paper*, 2013, 19275.
- Ariely, Dan and Klaus Wertenbroch**, “Procrastination, deadlines, and performance: Self-control by precommitment,” *Psychological Science*, 2002, 13 (3), 219–224.
- Banerjee, Abhijit V, Rukmini Banerji, Esther Duflo, Rachel Glennerster, and Stuti Khemani**, “Pitfalls of Participatory Programs: Evidence from a randomized evaluation in education in India,” *American Economic Journal: Economic Policy*, 2010, pp. 1–30.
- Baumeister, Roy F**, “Choking under pressure: self-consciousness and paradoxical effects of incentives on skillful performance,” *Journal of Personality and Social Psychology*, 1984, 46 (3), 610.
- Bettinger, Eric P**, “The effect of charter schools on charter students and public schools,” *Economics of Education Review*, 2005, 24 (2), 133–147.
- , “Paying to learn: The effect of financial incentives on elementary school test scores,” *Review of Economics and Statistics*, 2012, 94 (3), 686–698.
- Blimpo, Moussa P**, “Team incentives for education in developing countries: A randomized field experiment in Benin,” *American Economic Journal: Applied Economics*, 2014, 6 (4), 90–109.
- Butler, JS, Douglas A Carr, Eugenia F Toma, and Ron Zimmer**, “Choice in a world of new school types,” *Journal of Policy Analysis and Management*, 2013, 32 (4), 785–806.
- Calabrese, Stephen M, Dennis N Epple, and Richard E Romano**, “Inefficiencies from metropolitan political and fiscal decentralization: Failures of Tiebout competition,” *The Review of Economic Studies*, 2012, 79 (3), 1081–1111.
- Cameron, A Colin, Jonah B Gelbach, and Douglas L Miller**, “Bootstrap-based improvements for inference with clustered errors,” *Review of Economics and Statistics*, 2008, 90 (3), 414–427.
- Chakrabarti, Rajashri**, “Can increasing private school participation and monetary loss in a voucher program affect public school performance? Evidence from Milwaukee,” *Journal of Public Economics*, 2008, 92 (5), 1371–1393.

- , “Do vouchers lead to sorting under random private-school selection? Evidence from the Milwaukee voucher program,” Technical Report, Staff Report, Federal Reserve Bank of New York 2009.
- Cohen, Peter A, James A Kulik, and Chen-Lin C Kulik**, “Educational outcomes of tutoring: A meta-analysis of findings,” *American educational research journal*, 1982, 19 (2), 237–248.
- Cortes, Kalena E and Lei Zhang**, “The Incentive Effects of the Top 10% Plan,” *Under Review*, 2011.
- Cullen, Julie B, Brian A Jacob, and Steven Levitt**, “The Effect of School Choice on Participants: Evidence from Randomized Lotteries,” *Econometrica*, 2006, 74 (5), 1191–1230.
- , **Mark C Long, and Randall Reback**, “Jockeying for position: Strategic high school choice under Texas’ top ten percent plan,” *Journal of Public Economics*, 2013, 97, 32–48.
- Deming, David J**, “Better schools, less crime?,” *The Quarterly Journal of Economics*, 2011, p. qjr036.
- , **Justine S Hastings, Thomas J Kane, and Douglas O Staiger**, “School Choice, School Quality, and Postsecondary Attainment,” *American Economic Review*, 2014, 104 (3), 991–1013.
- Ding, Weili and Steven F Lehrer**, “Do peers affect student achievement in China’s secondary schools?,” *The Review of Economics and Statistics*, 2007, 89 (2), 300–312.
- Dobbie, Will and Roland G Fryer**, “Are high-quality schools enough to increase achievement among the poor? Evidence from the Harlem Children’s Zone,” *American Economic Journal: Applied Economics*, 2011, pp. 158–187.
- and – , “Getting beneath the Veil of Effective Schools: Evidence from New York City,” *American Economic Journal: Applied Economics*, 2013, 5 (4), 28–60.
- Epple, Dennis and Richard E Romano**, “Competition between private and public schools, vouchers, and peer-group effects,” *American Economic Review*, 1998, pp. 33–62.
- and **Richard Romano**, “Peer effects in education: A survey of the theory and evidence,” *Handbook of social economics*, 2011, 1 (11), 1053–1163.
- Fantuzzo, John W, Judith A King, and Lauren R Heller**, “Effects of reciprocal peer tutoring on mathematics and school adjustment: A component analysis,” *Journal of Educational Psychology*, 1992, 84 (3), 331.

- Ferrari, Joseph R, Judith L Johnson, and William George McCown**, *Procrastination and task avoidance: Theory, research, and treatment*, Springer, 1995.
- Frederick, Shane, George Loewenstein, and Ted O'donoghue**, "Time discounting and time preference: A critical review," *Journal of Economic Literature*, 2002, 40 (2), 351–401.
- Fryer, Roland G**, "Financial Incentives and Student Achievement: Evidence from Randomized Trials\*," *The Quarterly journal of economics*, 2011, 126 (4), 1755–1798.
- Fuchs, Douglas, Lynn S Fuchs, Patricia G Mathes, and Deborah C Simmons**, "Peer-assisted learning strategies: Making classrooms more responsive to diversity," *American Educational Research Journal*, 1997, 34 (1), 174–206.
- Galiani, Sebastian, Paul Gertler, and Ernesto Schargrodsky**, "School decentralization: Helping the good get better, but leaving the poor behind," *Journal of Public Economics*, 2008, 92 (10), 2106–2120.
- Greenwood, Charles R, Joseph C Delquadri, and R Vance Hall**, "Longitudinal effects of classwide peer tutoring.," *Journal of educational psychology*, 1989, 81 (3), 371.
- Hastings, Justine S and Jeffrey M Weinstein**, "Information, School Choice, and Academic Achievement: Evidence from Two Experiments\*," *The Quarterly journal of economics*, 2008, 123 (4), 1373–1414.
- , **Thomas J Kane, and Douglas O Staiger**, "Heterogeneous preferences and the efficacy of public school choice," *unpublished working paper*, 2008.
- Hoxby, Caroline M**, "Does competition among public schools benefit students and taxpayers?," *American Economic Review*, 2000, pp. 1209–1238.
- and **Christopher Avery**, "The Missing " One-Offs": The Hidden Supply of High-Achieving, Low-Income Students," *Brookings Papers on Economic Activity*, 2013, 2013 (1), 1–65.
- Hsieh, Chang-Tai and Miguel Urquiola**, "The effects of generalized school choice on achievement and stratification: Evidence from Chile's voucher program," *Journal of public Economics*, 2006, 90 (8), 1477–1503.
- Imbens, Guido W and Joshua D Angrist**, "Identification and estimation of local average treatment effects," *Econometrica: Journal of the Econometric Society*, 1994, pp. 467–475.

- Karau, Steven J and Janice R Kelly**, “The effects of time scarcity and time abundance on group performance quality and interaction process,” *Journal of Experimental Social Psychology*, 1992, 28 (6), 542–571.
- Kremer, Michael and Alaka Holla**, “Improving education in the developing world: what have we learned from randomized evaluations?,” *Annual Review of Economics*, 2009, 1, 513.
- , **Edward Miguel, and Rebecca Thornton**, “Incentives to learn,” *The Review of Economics and Statistics*, 2009, 91 (3), 437–456.
- Krueger, Alan B and Pei Zhu**, “Inefficiency, Subsample Selection Bias, and Nonrobustness A Response to Paul E. Peterson and William G. Howell,” *American Behavioral Scientist*, 2004, 47 (5), 718–728.
- Levin, Henry M**, “Educational vouchers: Effectiveness, choice, and costs,” *Journal of Policy Analysis and management*, 1998, 17 (3), 373–392.
- **et al.**, “Cost-Effectiveness of Four Educational Interventions.,” 1984.
- Li, Hongbin, Lingsheng Meng, Xinzheng Shi, and Binzhen Wu**, “Does attending elite colleges pay in China?,” *Journal of Comparative Economics*, 2012, 40 (1), 78–88.
- Li, Tao, Li Han, Linxiu Zhang, and Scott Rozelle**, “Encouraging classroom peer interactions: Evidence from Chinese migrant schools,” *Journal of Public Economics*, 2014, 111, 29–45.
- Long, Mark C**, “Race and College Admissions: An Alternative to Affirmative Action?,” *Review of Economics and Statistics*, 2004, 86 (4), 1020–1033.
- , **Victor Saenz, and Marta Tienda**, “Policy Transparency and College Enrollment: Did the Texas Top Ten Percent Law Broaden Access to the Public Flagships?,” *The ANNALS of the American Academy of Political and Social Science*, 2010, 627 (1), 82–105.
- MacLeod, W Bentley and Miguel Urquiola**, “Competition and Educational Productivity: Incentives Writ Large.”
- Mizala, Alejandra and Miguel Urquiola**, “School markets: The impact of information approximating schools’ effectiveness,” *Journal of Development Economics*, 2013, 103 (C), 313–335.
- Muralidharan, Karthik and Venkatesh Sundararaman**, “The Aggregate Effect of School Choice: Evidence from a two-stage experiment in India,” 2013.

- O'Donoghue, Ted and Matthew Rabin**, "Choice and procrastination," *The Quarterly Journal of Economics*, 2001, 116 (1), 121–160.
- Peterson, Paul E, David Myers, and William G Howell**, "An Evaluation of the New York City School Choice Scholarships Program: The First Year," 1998.
- Rohrbeck, Cynthia A, Marika D Ginsburg-Block, John W Fantuzzo, and Traci R Miller**, "Peer-assisted learning interventions with elementary school students: A meta-analytic review," *Journal of Educational Psychology*, 2003, 95 (2), 240.
- Rothstein, Jesse**, "Does Competition among Public Schools Benefit Students and Taxpayers? Comment," *The American Economic Review*, 2007, pp. 2026–2037.
- Rouse, Cecilia E**, "Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program," *Quarterly Journal of Economics*, 1998, pp. 553–602.
- **and Lisa Barrow**, "School vouchers and student achievement: Recent evidence and remaining questions," *Annual Review of Economics*, 2009, 1 (1), 17–42.
- Steel, Piers**, "The nature of procrastination: a meta-analytic and theoretical review of quintessential self-regulatory failure," *Psychological Bulletin*, 2007, 133 (1), 65.
- Tice, Dianne M and Roy F Baumeister**, "Longitudinal study of procrastination, performance, stress, and health: The costs and benefits of dawdling," *Psychological Science*, 1997, pp. 454–458.
- Topping, Keith J**, "Trends in peer learning," *Educational psychology*, 2005, 25 (6), 631–645.
- Walters, Christopher**, "A structural model of charter school choice and academic achievement," *Unpublished working paper, UC-Berkeley*, 2013.
- Webb, Noreen M and Sydney Farivar**, "Promoting helping behavior in cooperative small groups in middle school mathematics," *American Educational Research Journal*, 1994, 31 (2), 369–395.
- Witte, John F**, "Achievement effects of the Milwaukee voucher program," in "American Economics Association Annual Meeting, New Orleans" 1997, pp. 4–6.
- Yerkes, Robert M and John D Dodson**, "The relation of strength of stimulus to rapidity of habit-formation," *Journal of comparative neurology and psychology*, 1908, 18 (5), 459–482.

**Yi, Hongmei, Linxiu Zhang, Renfu Luo, Yaojiang Shi, Di Mo, Xinxin Chen, Carl Brinton, and Scott Rozelle**, “Dropping out: Why are students leaving junior high in China’s poor rural areas?,” *International Journal of Educational Development*, 2012, 32 (4), 555–563.

**Zhang, Hongliang**, “The Mirage of Elite Schools: Evidence from Lottery-based School Admissions in China’,” *Chinese University of Hong Kong*, 2012.